Université Catholique de l'Afrique de l'Ouest
Faculté des Sciences de Gestion
INSTITUT SUPERIEUR DE GESTION SAINT MICHEL
Science – Foi – Action

Agrément : n° 05/AG/SAC/MESUCURRS/DES/DFS

Habilitation : N° RepSEN/Ensup-priv/HA/015-2017

SPECIALITE : INFORMATIQUE DE GESTION

MEMOIRE

Présenté par

M Mouhamadou DIABY

Pour l'obtention du diplôme de

Master 2 en INFORMATIQUE DE GESTION

SUJET

ETUDE ET CONCEPTION D'UN SYSTEME D'INFORMATION DECISIONNELLE POUR LA GESTION DE L'ENERGIE DANS LE DOMAINE INDUSTRIEL DE DIAMNIADIO : CAS DE L'APROSI

Soutenu à UCAO/Saint Michel le 30/03/2019 devant le jury composé de :

Président : Pr Cheikh Amadou Bamba GUEYE	Professeur Titulaire en Informatique	UCAD
Directeur de mémoire : Pr Samba NDIAYE	Professeur Titulaire en Informatique	UCAD
Co-encadreur : Dr Edouard Ngor SARR	Assistant en Informatique	UCAO
Examinateur : Dr Reine Marie MARONE	Assistante en Informatique	UCAO

Année 2017-2018

DÉDICACE

- A l'âme de mon grand frère Mody Diaby (1979-2010)
- A mes parents qui ont fait beaucoup de sacrifice en France comme au Sénégal pour m'inculquer des valeurs.
- A ma femme Aissatou Dieye pour tout son engagement a mes cotes.
- A ma fille Sandia Diaby que je souhaite longue vie et sante
- A ma petite sœur que je souhaite une bonne continuation dans ses projets
- A mon cousin, ami et partenaire professionnel Sourakhata Tirera

REMERCIEMENTS

Mes remerciements à :

- Dieux le tout puissant qui nous a donné la volonté, le courage, la force et les moyens pour réussir et donner le meilleur de nous-mêmes.
- mon encadrant Monsieur EDOUARD SARR d'avoir accepté de superviser ce travail et pour son savoir-faire qui m'ont été d'une grande aide.
- aux professeurs de l'UCAO qui doivent voir dans ce travail la fierté d'un savoir bien acquis.
- les membres du jury qui nous font l'honneur d'évaluer ce travail.
- les dirigeants de l'APROSI [15]. a travers son Secrétaire General Monsieur BOHOUM SOW ainsi que Monsieur OUMAR NDIAYE et Monsieur BAMBA THIAM de la SENELEC qui ont contribué à l'accomplissement de ce travail.

RÉSUMÉ

Les nouvelles technologies de l'information de nos jours sont utilisés par les décideurs et dirigeants des entreprises pour obtenir une connaissance approfondie de l'entreprise et de définir leurs stratégies d'affaires, dans le but d'acquérir un avantage concurrentiel, d'améliorer la performance de l'entreprise, de répondre plus rapidement aux changements, d'augmenter la rentabilité, et d'une façon générale la création de valeur ajoutée de l'entreprise. Avec l'apparition des Data Warehouse, tous les utilisateurs peuvent désormais accéder facilement à l'information stratégique, ceci permet à l'entreprise d'être plus réactive dans sa prise de décisions. Ce mémoire traite l'étude et la conception d'un Data Warehouse pour le suivi de la consommation d'énergie dans le Domaine Industriel de Diamniadio, afin de permettre à l'APROSI [15]. de faire des analyses sur les résultats et prendre des décisions. L'approche retenue pour la réalisation de notre Data Warehouse est divisée en quatre phases principales: dans une première phase, nous avons effectué l'identification des besoins, une étude des outils techniques, ensuite une étude sur des différents outils d'implémentations d'entrepôt de données (SQLServer, Oracle, Pentaho, Talend Open Studio) et à la fin la conception du Data Warehouse [14].

Mots-clés : Data Warehouse, Diamniadio, ETL

ABSTRACT

New information technologies nowadays are used by decision-makers and business leaders to gain in-depth knowledge of the business and define their business strategies, in order to gain a competitive edge, improve the company's performance, to respond more quickly to changes, to increase profitability, and in a general way the creation of added value of the company. With the advent of the Data Warehouse, all users can now easily access strategic information, allowing the company to be more responsive in making decisions. This thesis deals with the study and design of a Data Warehouse for the monitoring of energy consumption in the Diamniadio Industrial Estate, in order to allow APROSI [15]. to analyze the results and make decisions. The approach adopted for the realization of our Data Warehouse is divided into four main phases: in a first phase, we carried out the identification of the needs, a study of the technical tools, then a study on various tools of implementations of data warehouse (SQLServer, Oracle, Pentaho [16], Talend Open Studio) and in the end the design of the Data Warehouse [14].

Keywords: Data Warehouse, Diamniadio, ETL

TABLES DES MATIÈRES

Dédicace i
Remerciements ii
Résumé iii
Abstract iv
Tables des matières v
Liste des figures viii
Liste des tableaux ix
Liste des abréviations x
CHAPITRE 1: CADRE DU PROJET 11
1. Introduction Générale 12
2. Contexte de l'étude 12
3. Problématique 13
4. Objectif de la recherche 15
5. Intérêt du sujet 15
6. Limitation de l'étude 16
7. Présentation de l'APROSI [15]. 17
8. Notre proposition 18
9. Méthodologie et Techniques de recherche 19
9.1. Phase de recueil des exigences 19
9.2. Phase de conception et de choix technique 19
10. Annonce du Plan 19
Chapitre 2 : Généralité de l'Informatique Décisionnelle 21
1. Informatique décisionnelle 22
2. Enjeux de l'informatique décisionnelle 22
3. Fonctions essentielles de l'informatique décisionnelle 24
3.1. Collecte 24
3.2. Intégration 25
3.3. Diffusion (ou distribution) 25
3.4. Présentation 26
3.5. Administration 26
4. Architecture des Systèmes d'Informations Décisionnelles 26

5. Caractéristique du Data Warehouse 28
6. Caractéristiques de l'ETL 29
CHAPITRE 3: CONCEPTION ER RELISATION 31
1. Conception 32
1.1. Source des données de consommation d'énergie 32
1.2. Modélisation avec UML 32
1.2.1. UML 32
1.2.2. Diagramme de cas d'utilisation 33
1.2.3. Diagramme de classe 34
2. Conception du schéma multidimensionnel 35
2.1. Spécification des besoins 36
2.2. Formalisation des besoins 36
2.3. Définition des faits 36
2.4. Définition des dimensions et enrichissement 37
2.5. Démarche ascendante 37
2.6. Détermination des faits 37
2.7. Détermination des dimensions 38
2.8. Définition de la dimension temporelle 39
3. Environnement et Outils 40
3.1. Environnement Technique 40
3.1.1. Architecture d'un capteur sans fil 41
3.1.2. Architecture de communication 42
3.1.3. Applications des Réseaux capteurs sans fil 42
3.1.4. Consommation d'énergie dans les Réseaux capteur sans fil 43
3.2. Environnement Logiciel 44
3.2.1. Pentaho Data Integration 45
3.2.2. Talend Open Studio 45
3.2.3. SQL Server management studio 45
3.2.4. SQL Server Integration Services SSIS 45
3.2.5. NetBeans 46
3.2.6. Power AMC 46
4. Réalisation 47
4.1. Etape de la realisation 47

4.2. Création Le modèle multidimensionnel avec SQL Server Management Studio 48
4.3. Création de cube OLAP avec SQL Business Intelligence Management studio 48
4.4. Création de la source de données 49
4.5. Création de la vue de source de données 50
4.6. Création du cube 52
5. Déploiement 53
5.1. Déploiement de Cube 53
5.2. Traitement de cube 54
5.3. Exploration du Cube Navigateur 55
7. Cout du projet 57
CONCLUSION GENERALE 58
BIBLIOGRAPHIE : 59
ANNEXE 1: Object Table 61

LISTE DES FIGURES

figure 1. Architecture système décisionnel 22
figure 2. Composition du Datawarehouse 24
figure 3. Architecture des Systèmes d'Informations Décisionnelles 27
figure 4. Caractéristiques du data warehouse 28
figure 5. Données orientées sujet 29
figure 6. Données non volatiles 29
figure 7. ETL 30
figure 8. Diagramme de cas d'utilisation 33
figure 9. Diagramme de classe 34
figure 10. schéma multidimensionnel en fonction des besoins 35
figure 11. Architecture d'un capteur 41
figure 12. Architecture de communication 42
figure 13. Les étapes de la conception 48
figure 14. Démarrage de Projet Analyses services 49
figure 15. Nouvelle source de données 50
figure 16. Schéma sélection des tables de vues 51
figure 17. Vues Source de données 51
figure 18. sélection des nouvelles dimensions 52
figure 19. Nouveau Cube 53
figure 20. Déploiement du Cube 53
figure 21. Exécution 54
figure 22. Fin de Traitement 54
figure 23. Exploration du Cube Navigateur 55
figure 24. Consommation d'énergie par Capteur 56
figure 25. Consommation d'énergie par Entrepôt 56
figure 26. Consommation d'énergie par Machine 57
figure 27. Coût du projet 57

LISTE DES TABLEAUX

Tableau 1. Requêtes-types 36
Tableau 2. Spécification des besoins 36

LISTE DES ABRÉVIATIONS

- **APROSI [15]. :** Agence d'Aménagement et de Promotion des Sites Industriels
- **APIX :** Agence nationale chargée de la Promotion de l'Investissement et des grands travaux
- **DID** : Domaine Industriel de Diamniadio
- **ZES :** Zones Economiques spéciales
- **UCAO** : Université Catholique de l'Afrique de l'Ouest
- **SGBD:** Système de Gestion de Base de Données
- **DW :** Data Warehouse
- **DM** : Data Mart
- **ETL :** Extract Transform Load
- **DSS** : Décision Support System
- **BI** : Business Intelligence
- **SID** : Système d'Information Décisionnel
- **SIO :** Système d'Information Opérationnel
- **OLAP :** Online Analytical Processing
- **R-OLAP**: Relational- Online Analytical Processing
- **M-OLAP:** Multidimensional-Online Analytical Processing
- **H-OLAP:** Hybrid-Online Analytical Processing
- **ODS :** Operational Data Store
- **ADC:** Analog to Digital Converters

CHAPITRE 1: CADRE DU PROJET

1. Introduction Générale

La croissance massive des données et des appareils a un impact profond sur les entreprises. Toutes les entreprises du monde entier ont pour défi de se transformer en organisation numérique. Si votre organisation est déjà engagée dans ce processus de transformation numérique, vous comprenez son potentiel pour améliorer chaque aspect de votre entreprise. Votre capacité à exploiter la valeur des données peut aider à définir le succès de votre organisation et à découvrir de nouvelles façons de créer et de capturer de la valeur et de s'étendre à de nouveaux segments de clientèle. Jamais les opportunités et les défis n'ont été aussi grands. La plupart des entreprises disposent de plusieurs silos d'informations verticaux. Différentes applications collectent des données qui se chevauchent, parfois en conflit [5].

Les données non structurées, qui peuvent inclure des données, sont généralement collectées sans modèle de données prédéfini, mais doivent être connectées au reste de vos données. Les bases de données constituent le fondement du provisionnement de ces données, et l'analyse de données est devenue l'outil par lequel ces données sont converties en informations exploitables. Quels que soient l'emplacement des données et leur mode de stockage, les professionnels de la base de données doivent respecter les objectifs de temps et de point de récupération de l'organisation [4, 11].

Le processus décisionnel est né d'un besoin exprimé par les entreprises à cause du volume important des données à manipuler et à analyser au quotidien. Le système traditionnel ne suffit plus pour pérenniser l'activité de gestion des données volumineux, car celui-ci convient bien aux applications gérant l'activité de l'entreprise, mais inadapté au Décisionnel. Le processus décisionnel apporte la solution au problème de la croissance continuelle des données provenant de différentes sources et de différents formats en les homogénéisant et en les organisant dans des Data Mart [17] où elles sont historiées, résumées et consolidées. Le volume de données des entrepôts est important et va de centaines de giga-octets à des téraoctets, voir même encore davantage de nos jours[13].

2. Contexte de l'étude

Le Sénégal connait depuis trois ans une croissance économique évolutive de 6,5% en 2014 et 2015, de 6.6 % en 2016, et les dernières prévisions de Standard & Poor's l'annoncent supérieures à 6,7% jusqu'en 2019, soit une moyenne quinquennale de

6.5%. Le développement économique des entreprises au Sénégal et l'accroissement du volume de données dans les systèmes d'information sont de nos jours une réalité à laquelle chaque entreprise qui évoluent dans un environnement difficile à appréhender de faire face aux défis des prises de décisions. Il en résulte que la collecte et le stockage de données ne peuvent plus suffire car la prise de décision stratégique est de plus en plus complexe face à la concurrence et l'augmentation du nombre de paramètres à prendre en compte [1]. Les nouvelles technologies de l'information permettent de concevoir des systèmes d'information performants et novateurs. Avec l'apparition des entrepôts de données (Data Warehouse) tous les acteurs du Domaines industriel de Diamniadio peuvent désormais accéder à l'information stratégique.

Comment dans le Domaine industriel de Diamniadio s'organise la prise de décision?

Nous répondrons à cette question en organisant notre réflexion en quatre temps. D'abord nous commencerons par affiner et préciser la problématique en délimitant le champ de la réflexion. Ensuite, nous examinerons les concepts du Data Warehouse [14]. Pour terminer nous illustrerons nos propos par l'exemple de logiciels proposant des solutions plus ou moins complètes.

3. Problématique

Certes, le problème de gestion des données énergétiques du Domaine industriel de Diamniadio reste le plus saillant, auquel les responsables (APROSI [15].) qui sont les promoteurs développeur public cherchent à trouver des solutions les plus adéquates possibles afin d'assurer l'organisation et la structuration des zones d'activités, entre autre la fiabilité dans le traitement, la diffusion et la recherche des informations dans le but d'éviter la lourdeur et l'encombrement dans cet environnement. Dans le contexte actuel, l'intégration d'un Système d'Information et de Gestion des Domaines industriel de Diamniadio, est le choix le plus rationnel [8]. Ce Choix tient compte de l'évolution technologique, de l'organisation des activités de l'entreprise et de l'immensité des opérations.

Cependant, l'implantation ou l'intégration (la mise en place d un tel environnement, d une plateforme, application) d'outils de gestion demande des dispositions techniques, scientifiquement mesurées et préalablement étudiées qui doivent être prises en compte pour une parfaite modélisation du système d'information des Domaines industriel de Diamniadio.

Ce travail requiert beaucoup plus de précautions si l'on veut satisfaire les besoins des utilisateurs et atteindre dans un temps record les objectifs. L'objectif est donc d'apporter aux décideurs d'une entreprise les moyens de répondre à ces questions.

En somme :

- **Du point de vu organisationnelle** nous devons savoir que l'évaluation de la structure est fondée sur cinq critères suivants :
  - **Programme explicite de gestion de l'énergie:** Existe-t-il une vision explicite qui oriente les activités de gestion de l'énergie réalisées sur le site? -A-t-on fixé des objectifs pour l'amélioration du rendement énergétique?
  - **Engagement clair de la direction:** La haute direction du site s'est-elle engagée à améliorer le rendement énergétique et quelle est la nature de cet engagement?
  - **Équipes de l'énergie mises sur pied et principaux intervenants inclus:** Existe-t-il une équipe de gestion de l'énergie ou un groupe organisationnel chargé de superviser le programme de gestion de l'énergie?
  - **Obligation de rendre compte du rendement énergétique incombant à la production et aux opérations** : Les cadres hiérarchiques, l'équipe des opérations et les opérateurs doivent-ils rendre compte du rendement énergétique à leurs supérieurs? Si le personnel de la production et des opérations est tenu responsable du rendement, a-t-il le pouvoir d'amorcer des changements dans le but d'améliorer le rendement?
- **Point de vu de la gestion des procédures** nous devons savoir que l'évaluation de la structure est fondée sur sept critères suivants :
  - **Examens réguliers du rendement:** L'examen du rendement énergétique est-il aligné sur la fréquence des rapports produits par les systèmes d'analyse des données?
  - **Mesures assignées et suivi effectué:** Des mesures sont-elles assignées aux employés dans le cadre du processus d'examen?
  - **Mesures incitatives visant le personnel:** incite-t-on le personnel à améliorer le rendement énergétique de l'équipement dont il est

responsable? Dans l'affirmative, quelle est la nature des mesures incitatives ou de la motivation?

- **Affectation de ressources suffisantes:** des ressources suffisantes ont-elles été affectées à l'exploitation d'un système de gestion en matière de temps de gestion?
- **Examens réguliers:** à quelle fréquence examine-t-on l'exploitation du système de gestion afin de s'assurer qu'il reste efficace?
- **Calendriers des opérations et de l'entretien:** Existe-t-il des calendriers des opérations et de l'entretien dans le but de réduire le rendement énergétique ?
- **Systèmes de production de rapports actifs**: Existe-t-il un système de production de rapports sur le gaspillage de l'énergie ?

Face aux études menées dans le système existant de l'APROSI [15]. nous sommes amenés à formuler et prendre pour sujet d'études « ETUDE ET CONCEPTION D'UN SYSTEME D'INFORMATION DECISIONNELLE POUR LA GESTION DE L'ENERGIE DANS LE DOMAINE INDUSTRIEL DE DIAMNIADIO».

4. Objectif de la recherche

Le Domaine industriel dans lequel l'entreprise va évoluer est confronté à des problèmes complexes comme la centralisation et le traitement des données de nature énergétique.

Notre objectif est de proposer une solution complète au traitement des informations permettant de concevoir, de modéliser et de manipuler des bases de données multidimensionnelles assurant l'intégrité et la cohérence des données avec une capacité d'exploitation efficace de grandes quantités d'information [3].

Ces travaux visent à proposer un modèle adapté aux besoins décisionnels s'adressant à des décideurs non informaticiens. Néanmoins, nous constatons le manque de modèles conceptuels permettant de faire abstraction des contraintes techniques de stockage de données et de mieux appréhender les besoins de ces décideurs.

5. Intérêt du sujet

Le choix et l'intérêt de ce sujet portent sur, d'abord il est question de répondre a nos inspirations d'approfondir la notion d'informatique décisionnelle pour essayer de les

appliquer dans la vie courante et professionnelle et ensuite de réaliser un outil de travail pour la gestion de l'énergie dans le Domaine industriel de Diamniadio Ainsi que l'obtention du diplôme de Master en Informatique de Gestion et la présentation d'un mémoire de fin d'études supérieures dans ce domaine supposent la réussite aux cours dispensés pendant tout le parcours au département informatique d'UCAO St Michel. Enfin, la réalisation de cette étude sera considérée comme modèle dans l'informatisation des opérations relatives à la Gestion des Domaines industriels du Sénégal. Elle servira également d'un recueil ou d'un référentiel pour les futurs techniciens dans ce domaine ainsi que pour les différents chercheurs qui s'y intéresseront [7].

Dans un contexte décisionnel, l'objet de nos travaux est de proposer un modèle conceptuel, un langage de manipulation et une méthode de conception des Data Warehouse (DW) prenant en compte l'intégrité des données. Le modèle que nous souhaitons proposer permet de représenter plusieurs centres d'intérêts. Afin de comprendre le niveau conceptuel de la modélisation d'un Data Warehouse, on va définir deux concepts : le concept de fait et le concept de dimension :

- **Concept de fait** : Un fait représente un sujet d'analyse. Il est constitué de plusieurs mesures relatives au sujet traité. Ces mesures sont numériques et généralement valorisées de façon continue.
- **Concept de dimensions** : La dimensions est le critère suivant lequel on souhaite évaluer, quantifier, qualifier le fait. Cette modélisation permet de faciliter la mise en œuvre de corrélations lors des analyses décisionnelles.

6. Limitation de l'étude

Délimiter une étude c'est de préciser le champ d'investigation ainsi que sa temporalité, qui sont deux éléments indispensables à sa contextualisation. La délimitation de notre étude permet non seulement la justification du choix et de l'intérêt du sujet, mais restreinte au projet informatique de «Gestion de l'énergie dans le Domaine industriel de Diamniadio». En effet sa généralisation à toute gestion des Domaines industriels ne devra être extrapolée à des similitudes avec beaucoup plus de prudence ; car, chaque organisation a des spécifications propres et particulières. Enfin, cette étude est également restreinte temporellement aux activités du Domaine industriel de Diamniadio, compte tenu des mutations ou mouvements au niveau des restructurations du cadre de recherche et du sous-ensemble représentatif [11]. Nous

projetterons largement nos études sur les DataWarehouse prévues pour le Domaine industriel de Diamniadio.

7. Présentation de l'APROSI [15].

APROSI [15].[1], une volonté de l'Etat pour compléter le dispositif industriel. L'absence d'espaces aménagés et de locaux industriels, faute d'un organe approprié à cet effet, est un des obstacles à la massification des investissements privés dans notre pays. C'est pour pallier cette lacune qu'il a été envisagé notamment, la mise en place de l'Agence d'Aménagement et de Promotion des Sites Industriels

- APROSI [15].-par décret 2000-981 du 7 octobre 2002, pour la prise en charge de l'aménagement du foncier industriel ainsi que de la réalisation de bâtiments industriels adéquats.
- L'APROSI [15]., qui vient ainsi compléter le dispositif de développement du secteur privé, a pour missions :
    - D'identifier, d'acquérir et de constituer des réserves foncières, dans le respect de la législation en vigueur, pour l'aménagement de sites industriels en liaison avec la Direction générale des Impôts et Domaines, la Direction de l'Urbanisme et les collectivités locales concernées;
    - De mieux structurer la forme d'appui de l'Etat aux entreprises industrielles en matière d'affectation de sites;
    - D'assurer la gestion foncière des sites par la cession ou la location à usage industriel;
    - D'assurer l'aménagement et la promotion d'espaces bâtis destinés à des activités industrielles;
    - D'assurer l'aménagement des espaces communs destinés à l'incubation et à l'essaimage d'activités industrielles à haute valeur ajoutée en privilégiant la proximité de plates-formes industrielles dynamiques, de centres d'enseignement supérieur ou de formation professionnelle;¬De procéder ou faire procéder à la réhabilitation et à l'aménagement de sites destinés à abriter harmonieusement des activités industrielles;
    - De s'assurer de la bonne gestion et la maintenance des sites;

[1] http://www.aprosi.sn/ visité le 21/01/2019

- De contribuer au développement d'une offre compétitive de production et de services des opérateurs installés sur le site;
- De promouvoir toute action pour la rentabilisation et l'occupation optimales des sites industriels;
- D'entreprendre toute démarche nécessaire à la réhabilitation et l'aménagement des sites.

8. Notre proposition

Dans le domaine industriel de Diamniadio, le volume de données énergétiques traitées croît rapidement avec le temps et l'augmentation des entreprises dans la zone. Une entreprise dispose de plusieurs options pour traiter ce flux de données :

- les données anciennes sont effacées et l'entreprise ne conserve que les données actives ou un historique récent ;
- les données sont stockées dans une base et l'entreprise n'envisage pas d'usage immédiat
- les données sont stockées au fur et à mesure qu'elles arrivent de manière cohérente pour qu'elles soient exploitables directement.

Notre projet décisionnel correspond à cette dernière option. Il s'agit de traiter les données et de les stocker de manière cohérente au fur et à mesure qu'elles se présentent. C'est pour cela que le projet décisionnel est un projet sans limite dans le temps. C'est-à-dire que dès que l'entreprise commence ce projet, elle ne s'arrête pas (sauf cas exceptionnel). Pour mener à bien ce projet décisionnel, il existe une multitude d'outils, chacun étant plus ou moins adapté à la taille de l'entreprise, à la structure des données existantes et au type d'analyse désiré.

- Un SGBD relationnels et d'autres systèmes vont contenir les données d'exploitation.
- Un ETL extrait les données pertinentes et les charge dans l'ODS Operational Data Store du Data Warehouse [14].
- Des données sont structurées seront dans le Data Warehouse [14].
- Des Data Mart [17] qui exploitent une technologie X-OLAP [18] sont mis à jour à partir du Data Warehouse [14].
- Des rapports sont générés sur ces données.

9. Méthodologie et Techniques de recherche

Pour rassembler les informations nécessaires à notre étude, nous utiliserons diverses techniques, celles relatives au fonctionnement du domaine de ***Gestion***. Ces informations nous permettront de comprendre les besoins des utilisateurs afin de préparer les solutions possibles ainsi que leurs appréciations [9].

Dans le cadre de notre travail, nous allons mettre en place une plateforme informatique qui permettra au décideur d'anticiper sur les décisions. Pour notre système nous aurons un fichier Excel que nous allons intégrer dans une base de données opérationnelles qui est comme base de données transactionnelles .Nous allons utiliser les datamining pour explorer les données de notre Data Mart [17] en vue de faire l'extraction.

9.1. Phase de recueil des exigences

Trois domaines doivent être particulièrement documentés :

- le type d'information dont l'utilisateur des rapports a besoin.
- le type de restitution (ergonomie, fréquence, vitesse de restitution)
- le système technique existant : technologies utilisées

9.2. Phase de conception et de choix technique

En fonction des exigences recueillies :

- quels sont les éléments de la chaîne de la valeur décisionnelle qui doivent être implémentés ?
- doit-on seulement créer un rapport sur un cube OLAP [18] existant ?
- construire toute la chaîne ?
- quelles sont précisément les données que l'on doit manipuler ?

Cela conduit au choix de technologies précises et à un modèle particulier.

10. Annonce du Plan

Afin de présenter notre travail, le présent mémoire est organisé en trois parties et se présente comme suit :

Chapitre 1 (Cadre du projet): Dans ce chapitre, nous présentons le projet, la société d'accueil, l'analyse de l'existant et la spécification des besoins.

Chapitre 2 (Généralité de l'Informatique Décisionnelle): Dans ce deuxième chapitre, nous développons les notions de base de l'informatique décisionnelle.

Chapitre 3 (Conception et Réalisation) : Nous présentons, dans ce chapitre, les différentes phases de la construction de l'entrepôt de données et le développement de l'ETL Dans un premier temps nous allons faire la Conception et Modélisation dans un second Les environnements et les outils utilises en fin la réalisation qui consiste a la mise en place du Data Warehouse [14].

Une conclusion Générale est proposée pour synthétiser le travail réalisé et de citer les perspectives du projet.

CHAPITRE 2 : GÉNÉRALITÉ DE L'INFORMATIQUE DÉCISIONNELLE

1. Informatique décisionnelle

L'informatique décisionnelle (en anglais : DSS pour Décision Support System ou encore BI pour Business Intelligence) désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données, matérielles ou immatérielles, d'une entreprise en vue d'offrir une aide à la décision et de permettre aux responsables de la stratégie d'entreprise d'avoir une vue d'ensemble de l'activité traitée. Ce type d'application utilise en règle générale un entrepôt de données (ou Data Warehouse en anglais) pour stocker des données transverses provenant de plusieurs sources hétérogènes et fait appel à des traitements par lots pour la collecte de ces informations.

Nous présentons dans un premier temps les enjeux de l'informatique décisionnelle. Nous abordons, ensuite, ses termes et les concepts clés en détaillant la notion d'ETL et d'entrepôt de données. Puis, nous développons les notions de Big Data et de bases de données NoSQL.

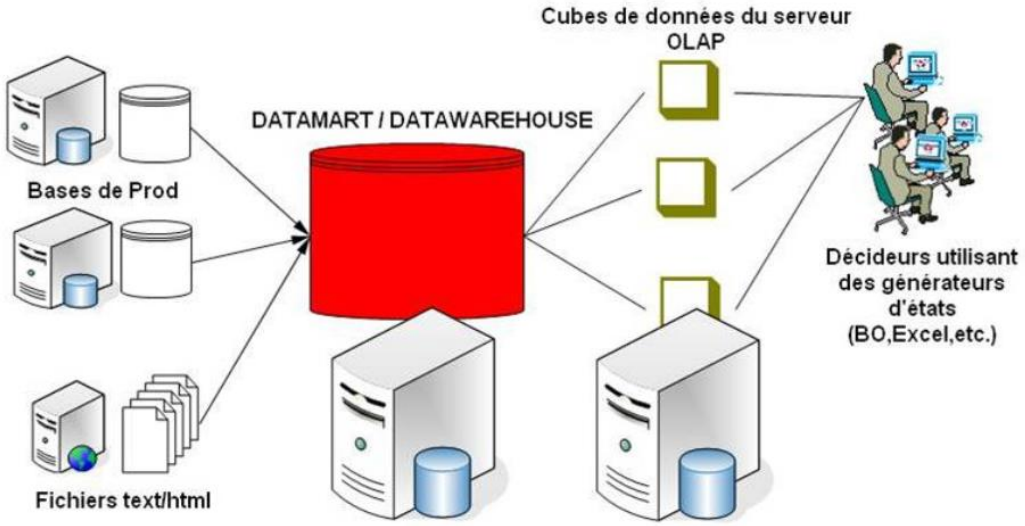

figure 1. Architecture système décisionnel

2. Enjeux de l'informatique décisionnelle

De nos jours, les données applicatives métier sont stockées dans une (ou plusieurs) base(s) de données relationnelle(s) ou non relationnelles. Ces données sont extraites, transformées et chargées dans un entrepôt de données généralement par un outil de type ETL (Extract-Transform-Load) ou en français ETC (Extraction-Transformation-Chargement).

Un entrepôt de données peut prendre la forme d'un Data Warehouse ou d'un Data Mart [17]. En règle générale, le Data Warehouse globalise toutes les données applicatives de l'entreprise, tandis que les datamarts (généralement alimentés depuis les données du Data Warehouse) sont des sous-ensembles d'informations concernant un métier particulier de l'entreprise (marketing, risque, contrôle de gestion, ...). Le terme comptoir de données ou magasin de données est aussi utilisé pour désigner un datamart.

Le reporting est probablement l'application la plus utilisée encore aujourd'hui de l'informatique décisionnelle, il permet aux gestionnaires :de sélectionner des données relatives à telle période, telle production, tel secteur de clientèle, etc., de trier, regrouper ou répartir ces données selon les critères de leur choix, de réaliser divers calculs (totaux, moyennes, écarts, comparatif d'une période à l'autre, ...), de présenter les résultats d'une manière synthétique ou détaillée, le plus souvent graphique selon leurs besoins ou les attentes des dirigeants de l'entreprise.

Les programmes utilisés pour le reporting permettent bien sûr de reproduire de période en période les mêmes sélections et les mêmes traitements et de faire varier certains critères pour affiner l'analyse. Mais le reporting n'est pas à proprement parler une application d'aide à la décision. L'avenir appartient plutôt aux instruments de type tableau de bord équipés de fonctions d'analyses multidimensionnelles de type OLAP [18]. Fonction OLAP qui peut être obtenue de différentes façons par exemple via une base de données relationnelle R-OLAP, ou multidimensionnelle M-OLAP, voire aussi en H-OLAP.

Les Data Mart [17] et/ou les Data Warehouses peuvent ainsi permettre via l'OLAP [18] l'analyse très approfondie de l'activité de l'entreprise, grâce à des statistiques recoupant des informations relatives à des activités apparemment très différentes ou très éloignées les unes des autres, mais dont l'étude fait souvent apparaître des dysfonctionnements, des corrélations ou des possibilités d'améliorations très sensibles. L'interopérabilité entre les systèmes d'entrepôt de données, les applications informatiques ou de gestion de contenu, et les systèmes de reporting est réalisée grâce à une gestion des métadonnées.

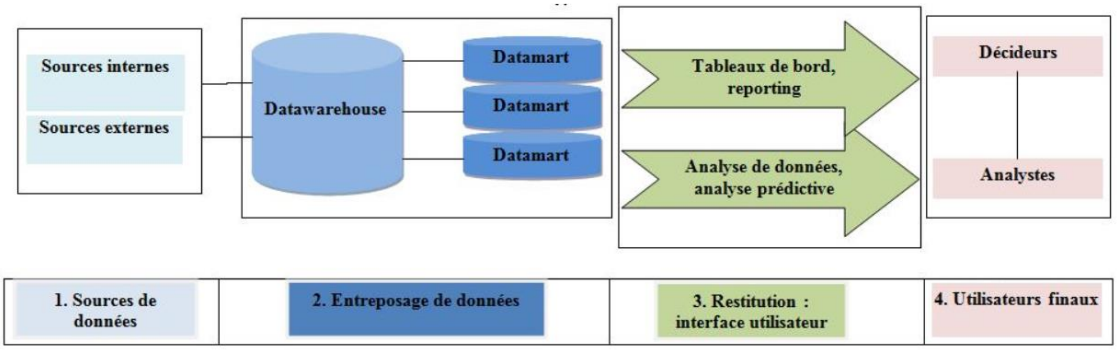

figure 2. Composition du Datawarehouse

3. Fonctions essentielles de l'informatique décisionnelle

Un système d'information décisionnel (SID) assure quatre fonctions fondamentales, à savoir la collecte, l'intégration, la diffusion et la présentation des données. À ces quatre fonctions s'ajoute une fonction de contrôle du SID lui-même, l'administration.

3.1. Collecte

La collecte (parfois appelée datapumping) est l'ensemble des tâches consistant à détecter, à sélectionner, à extraire et à filtrer les données brutes issues des environnements pertinents compte tenu du périmètre du SID. Les sources de données internes et/ou externes étant souvent hétérogènes tant sur le plan technique que sur le plan sémantique, cette fonction est la plus délicate à mettre en place dans un système décisionnel complexe. Elle s'appuie notamment sur des outils d'ETL (extract-transform-load pour extraction-transformation-chargement).

Cette alimentation utilise les données sources issues des systèmes transactionnels de production, le plus souvent sous forme de: compte-rendu d'événement ou compte-rendu d'opération : c'est le constat au fil du temps des opérations (achats, ventes, écritures comptables, ...), le film de l'activité de l'entreprise compte-rendu d'inventaire ou compte-rendu de stock : c'est l'image photo prise à un instant donné (à une fin de période : mois, trimestre, ...) de l'ensemble du stock (les clients, les contrats, les commandes, les encours, ...).

La fonction de collecte joue également, au besoin, un rôle de recodage. Une donnée représentée différemment d'une source à une autre impose le choix d'une représentation unique pour les futures analyses.

3.2. Intégration

L'intégration consiste à concentrer les données collectées dans un espace unifié, dont le socle informatique essentiel est l'entrepôt de données. Élément central du dispositif, il permet aux applications décisionnelles de bénéficier d'une source d'information commune, homogène, normalisée et fiable, susceptible de masquer la diversité de l'origine des données. Au passage les données sont épurées ou transformées par :

- Un filtrage et une validation des données en vue du maintien de la cohérence d'ensemble (les valeurs acceptées par les filtres de la fonction de collecte mais susceptibles d'introduire des incohérences de référentiel par rapport aux autres données doivent être soit rejetées, soit intégrées avec un statut spécial)
- Une synchronisation (s'il y a nécessité d'intégrer en même temps ou à la même « date de valeur » des événements reçus ou constatés de manière décalée ou déphasée)
- Une certification (pour rapprocher les données de l'entrepôt des autres systèmes « légaux » de l'entreprise comme la comptabilité ou les déclarations réglementaires). C'est également dans cette fonction que sont effectués éventuellement les calculs et les agrégations (cumuls) communs à l'ensemble du projet.
- La fonction d'intégration est généralement assurée par la gestion de métadonnées, qui assurent l'interopérabilité entre toutes les ressources informatiques, que ce soit des données structurées (bases de données accédées par des progiciels ou applications), ou des données non structurées (documents et autres ressources non structurées, manipulés par les systèmes de gestion de contenu).

3.3. Diffusion (ou distribution)

La diffusion met les données à la disposition des utilisateurs, selon des schémas correspondant au profil ou au métier de chacun, sachant que l'accès direct à l'entrepôt de données ne correspondrait généralement pas aux besoins d'un décideur ou d'un analyste. L'objectif prioritaire est de segmenter les données en contextes informationnels fortement cohérents, simples à utiliser et correspondant à une activité décisionnelle particulière. Alors qu'un entrepôt de données peut héberger des centaines

ou des milliers de variables ou indicateurs, un contexte de diffusion raisonnable n'en présente que quelques dizaines au maximum.

Chaque contexte peut correspondre à un Data Mart [17], bien qu'il n'y ait pas de règles générales concernant le stockage physique. Très souvent, un contexte de diffusion est multidimensionnel, c'est-à-dire modélisable sous la forme d'un hypercube; il peut alors être mis à disposition à l'aide d'un outil OLAP [18]. Les différents contextes d'un même système décisionnel n'ont pas tous besoin du même niveau de détail. De nombreux agrégats ou cumuls, n'intéressant que certaines applications et n'ayant donc pas lieu d'être gérés en tant qu'agrégats communs par la fonction d'intégration, relèvent donc de la diffusion. Ces agrégats peuvent être, au choix, stockés de manière persistante ou calculés dynamiquement à la demande.

3.4. Présentation

Cette quatrième fonction, la plus visible pour l'utilisateur, régit les conditions d'accès de l'utilisateur aux informations. Elle assure le fonctionnement du poste de travail, le contrôle d'accès, la prise en charge des requêtes, la visualisation des résultats sous une forme ou une autre. Elle utilise toutes les techniques de communication possibles (outils bureautiques, requêteurs et générateurs d'états spécialisés, infrastructure web, télécommunications mobiles, etc).

3.5. Administration

C'est la fonction transversale qui supervise la bonne exécution de toutes les autres. Elle pilote le processus de mise à jour des données, la documentation sur les données (les méta-données), la sécurité, les sauvegardes, la gestion des incidents.

4. Architecture des Systèmes d'Informations Décisionnelles

Les premiers SID conçus au sein des entreprises étaient des systèmes avec très peu d'utilisateurs, élaborés parfois par les utilisateurs eux-mêmes, exploitant le plus souvent les données d'un seul SIO. Ces systèmes étaient la plupart du temps composés uniquement d'un extracteur de données et d'une petite base de données ou d'un tableur. Comme les SIO avant l'arrivée des PGI, les premiers SID furent développés au coup par coup, en dehors de toute démarche d'urbanisation.

Puis, au début des années 1990, les concepts de **Data Warehouse** (DW) et de **Data Mart [17]** (DM) firent leur apparition, popularisés par des auteurs tels que **Bill Inmon**

et Ralph Kimball. Aujourd'hui, selon l'architecture que l'on retiendra, le *data warehouse* et/ou les *Data Mart [17]* forment le cœur de tout système d'information décisionnel digne de ce nom.

Il n'existe pas d'architecture standard unique pour un SID, aucun consensus n'ayant jusqu'à présent émergé sur ce que pouvait être l'architecture idéale. Il n'existe pas de définition unique de ce qu'est un Data Warehouse ou un Data Mart [17], certains auteurs utilisant le terme data warehouse pour désigner ce que d'autres appellent un Data Mart [17], et réciproquement.

De même, la technique de modélisation multidimensionnelle ne joue pas toujours un rôle de la même importance en fonction de l'architecture implémentée : nous allons voir dans l'architecture proposée par Ralph Kimball, elle joue un rôle fondamental.

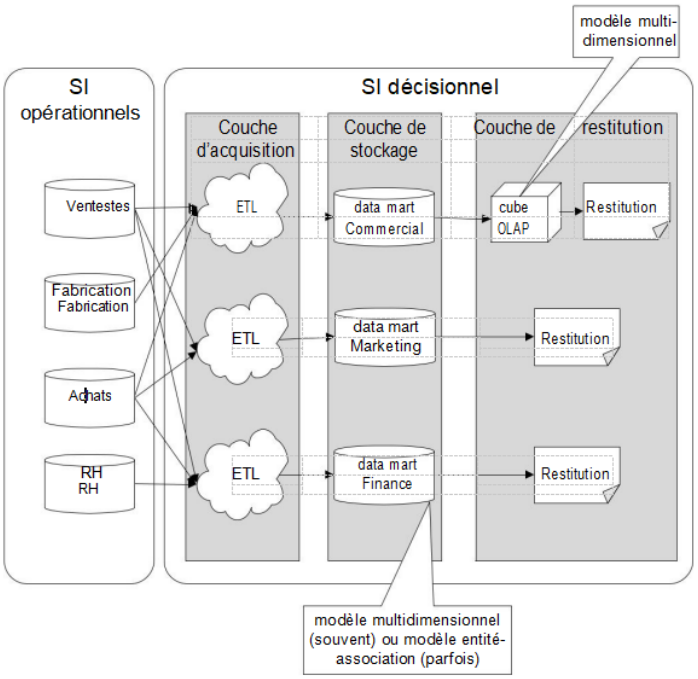

figure 3. Architecture des Systèmes d'Informations Décisionnelles

On peut classer les architectures des SID en trois1 grands types :

- l'architecture **data marts en silo**,
- l'architecture **corporate information factory** ou **enterprise data warehouse** (B. Inmon),
- l'architecture **dimensional data warehouse** ou **bus architecture** (R. Kimball).

5. Caractéristique du Data Warehouse

W. H. Inmon (1996): «Le data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision». Principe: mettre en place une base de données utilisée à des fins d'analyse

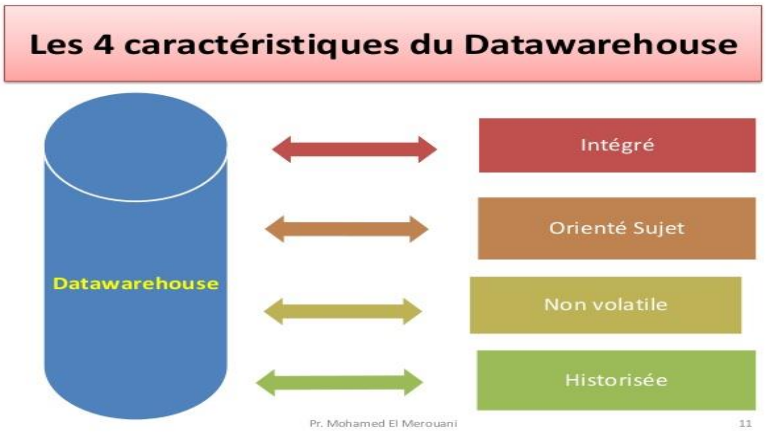

figure 4. Caractéristiques du data warehouse

Données orientées sujet: Les données d'un data Warehouse sont organisées par thèmes ou par sujets (Exemple:Pro duction,Vente, Marketingetc). Cette organisation p ermet de rassembler toutes les informations relatives à un thème précis afin de faciliter la prise de décision. Regroupe les informations des différents métiers. Ne tiens pas compte de l'organisation fonctionnelle des données.

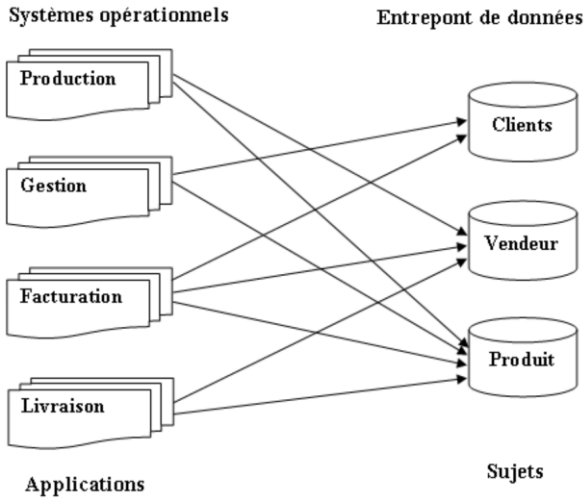

figure 5. Données orientées sujet

Données intégrées: Les données d'un Data warehouse proviennent de différentes sources hétérogènes. Cette hétérogénéité donne lieu à des conflits syntaxiques et sémantiques (conflits de représentation, de noms, de contexte et de mesure) entre les données des sources. L'intégration des données permet d'éliminer l'ensemble de ces conflits afin d'avoir une représentation uniforme et cohérente des données lors de leur chargement au niveau du Data Warehouse [14].

Données non volatiles: Les données d'un entrepôt sont généralement utilisées en mo de consultation. Elles peuvent être interrogées mais ne sont ni modifiées, ni supprimées (sauf dans les cas de rafraichissement de l'entrepôt).Ceci permet de conserver la traçabilité des informations afin de pouvoir effectuer des analyses sur une longue période

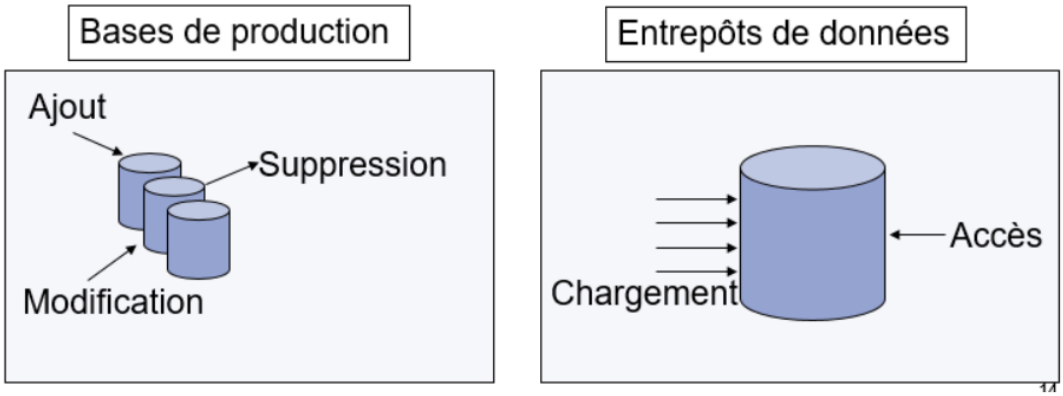

figure 6. Données non volatiles

Données historiées: La prise en compte de l'évolution des données est essentielle pour la prise de décision qui, par exemple, utilise des techniques de prédiction en s'appuyant sur les évolutions passées pour prévoir les évolutions futures.

6. Caractéristiques de l'ETL

L'acronyme « ETL » appartient à l'univers du Data Marketing. Avant de vous présenter les différents types d'ETL du marché, il n'est pas inutile de prendre le temps de définir clairement en quoi consiste un ETL. Par où commencer ? Partons des trois lettres qui composent cet acronyme :

E, pour Extract (Extraire).

T, pour Transform (Transformer).

L, pour Load (Charger).

Cet acronyme, en un sens, dit déjà tout ! L'ETL est un process par lequel « on » extrait dans un premier temps les données, on les transforme et, enfin, on les charge. Cela ne nous fait pas beaucoup avancer. Ne vous inquiétez pas, vous allez y voir plus clair dans quelques lignes.

Au-delà, ETL désigne, par métonymie, les *outils* permettant de réaliser ce process. C'est pour cette raison que notre article est consacré aux outils ETL, aux outils qui permettent de mettre en place un process ETL. Voici une manière de représenter la place et le rôle d'un ETL dans une architecture IT. Le schéma est proposé par Informatica, l'un des leaders (voire LE leader) du marché des ETL on-promise :

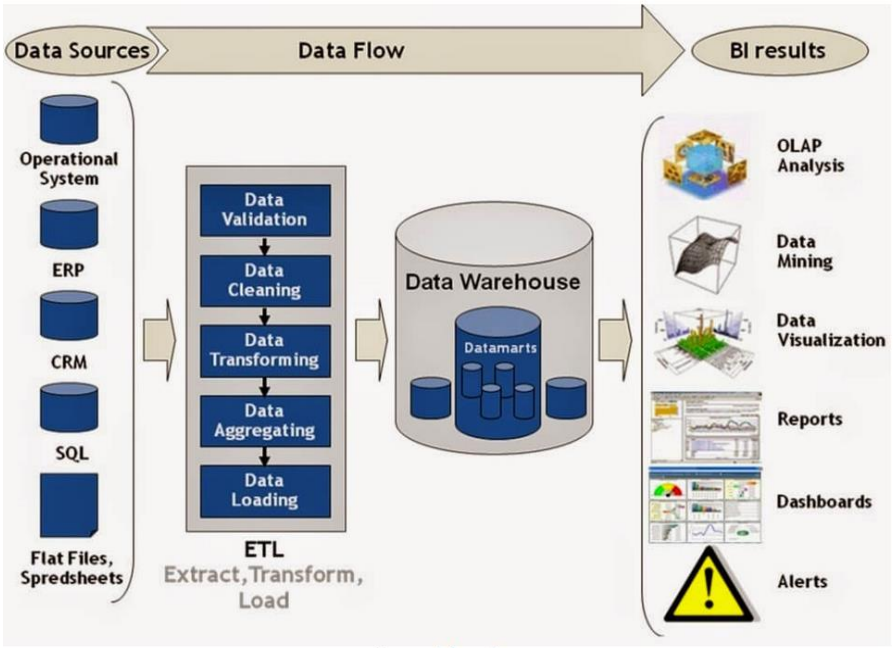

figure 7. ETL

CHAPITRE 3: CONCEPTION ER RELISATION

1. Conception

1.1. Source des données de consommation d'énergie

Cette partie constitue le niveau de l'organisation où le rendement de l'énergie est géré. Il n'y a pas de description précise de ce à quoi un centre de comptabilisation de l'énergie doit ressembler. Cela est dicté par le type de processus de production, l'importance des flux d'énergie et l'attribution des responsabilités au sein de l'organisation. Une source de données peut être une composante individuelle de l'équipement du processus, un secteur ou un processus de production ou toute une chaîne de production, Dans un premier temps nous présentons notre source de données qui défini la source globale. Notre source est composée de six tables qui sont reliées entre elles. Cette source des données de l'entreprise permet de suivre les différentes mesures générées par les capteurs et les sources d'où quelles proviennent.

1.2. Modélisation avec UML

1.2.1. UML

UML, Langage de Modélisation Unifié, de l'anglais Unified Modeling Language, est un langage de modélisation graphique à base de pictogrammes conçu pour fournir une méthode normalisée pour visualiser la conception d'un système. Il est couramment utilisé en développement logiciel et en conception orientée objet. La description de la programmation par objets a fait ressortir l'étendue du travail conceptuel nécessaire : définition des classes, de leurs relations, des attributs et méthodes, des interfaces, etc. Pour programmer une application, il ne convient pas de se lancer tête baissée dans l'écriture du code : il faut d'abord organiser ses idées, les documenter, puis organiser la réalisation en définissant les modules et étapes de la réalisation. C'est cette démarche antérieure à l'écriture que l'on appelle modélisation ; son produit est un modèle. Les spécifications fournies par la maîtrise d'ouvrage en programmation impérative étaient souvent floues : les articulations conceptuelles (structures de données, algorithmes de traitement) s'exprimant dans le vocabulaire de l'informatique, le modèle devait souvent être élaboré par celle-ci. L'approche objet permet en principe à la maîtrise d'ouvrage de s'exprimer de façon précise selon un vocabulaire qui, tout en transcrivant les besoins du métier, pourra être immédiatement compris par les informaticiens. En principe

seulement, car la modélisation demande aux maîtrises d'ouvrage une compétence et un professionnalisme qui ne sont pas aujourd'hui répandus.

1.2.2. Diagramme de cas d'utilisation

Nous cherchons dans ce diagramme à :

- Identifier les interactions du système avec son environnement : humains et autres systèmes.
- Identifier les besoins : fonctionnalités du systeme.
- Identifier les dépendances entre les fonctionnalités.

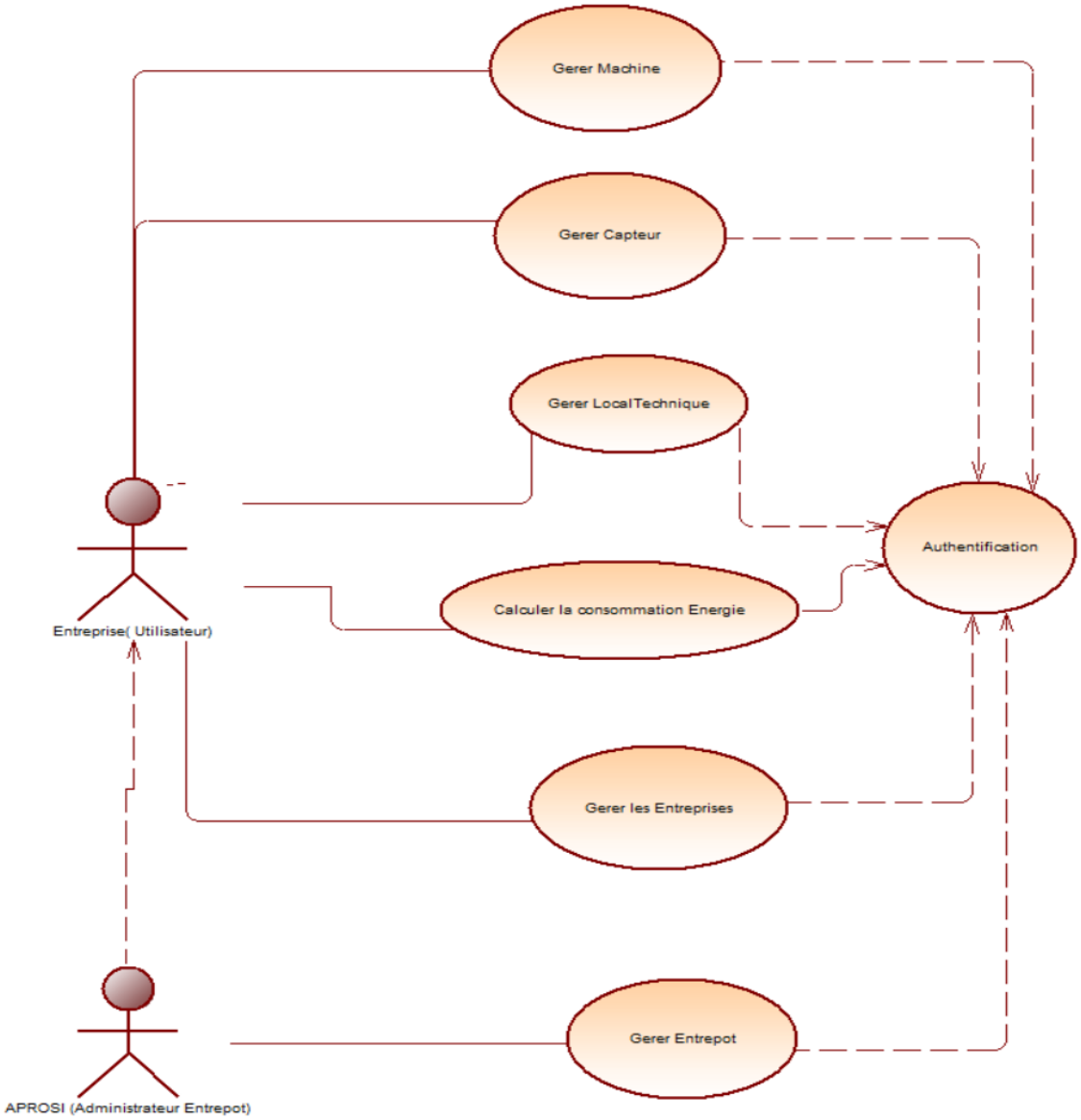

figure 8. Diagramme de cas d'utilisation

Ce diagramme illustre bien le fonctionnement de l'application, en effet on a deux types d'utilisateur à savoir l'utilisateur normal (les entreprises) et l'administrateur des entrepots (APROSI [15].) qui hérite toutes les actions de l'utilisateur et a en plus la tache de gestion des utilisateurs. Chaque action nécessite une authentification, qui se fait au démarrage de l'application. Le cas d'utilisation «calculer la consommation d'energie» peut être détaillé un peu plus car pour réaliser le calcul il faut passer par deux étapes, la première consiste a récupérer les données de consommation depuis les trendlogs et la deuxième consiste au calcul effectif et a l'enregistrement de cette valeur dans la base centrale.

1.2.3. Diagramme de classe

Ce diagramme de classes permet de représenter sous une forme simplifiée l'interaction des objets d'un système que vous êtes en train de modéliser. Les diagrammes de classes expriment la structure statique d'un système en termes de classes et de relations entre ces classes. C'est ce type de diagramme que nous allons créer dans ce tutoriel pour vous introduire aux Modèles Orienté Objets (MOO).

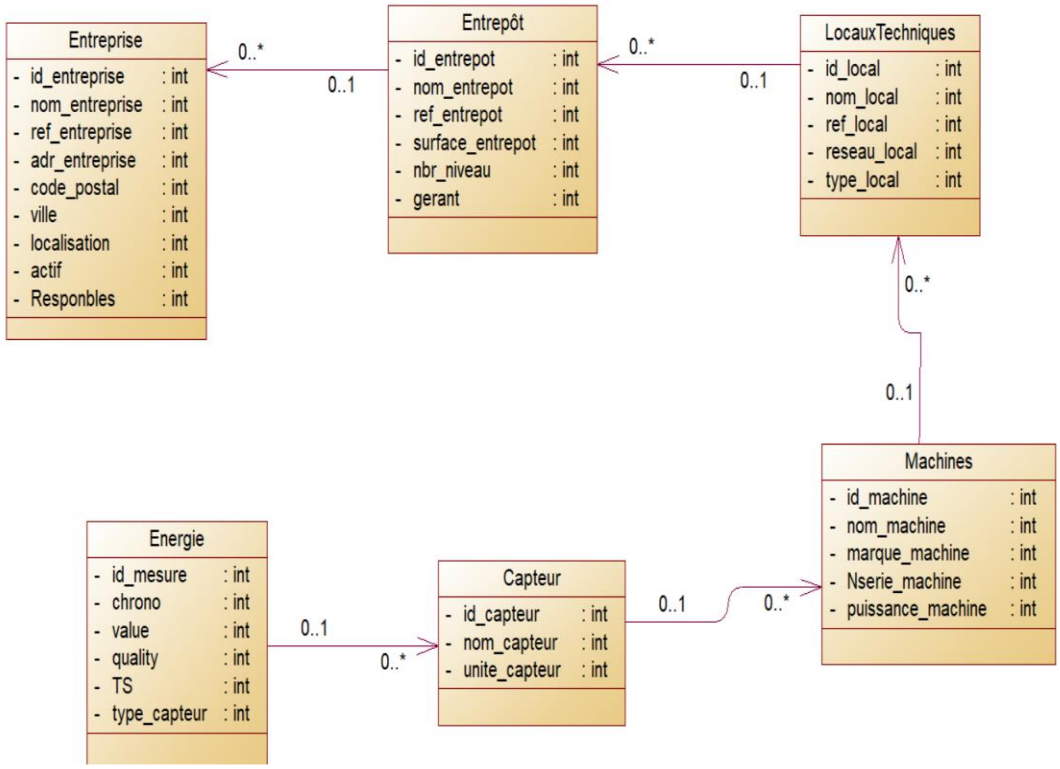

figure 9. Diagramme de classe

2. Conception du schéma multidimensionnel

Notre objectif est de spécifier le schéma de la base de données multidimensionnelle intégrant toutes les données d'aide à la prise de décision qui peuvent être extraites à partir de la source ou bien à partir des besoins des décideurs.

Pour répondre à ce besoin, nous avons suivi une approche mixte intégrant la description des besoins des décideurs à l'aide d'une démarche descendante et la description des données sources à l'aide d'une démarche ascendante

L'objectif de cette démarche est de concevoir un schéma multidimensionnel à partir des besoins des décideurs et des règles de gestion relatives aux données décisionnelles. Elle comporte trois étapes ; (1) la collecte des données, (2) la spécification des besoins et (3) la formalisation des besoins.

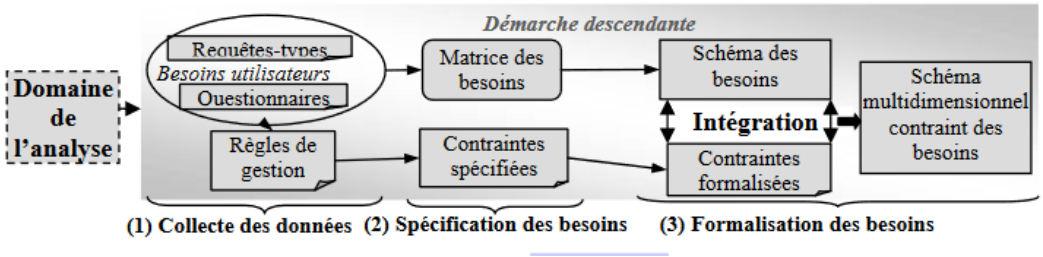

figure 10. schéma multidimensionnel en fonction des besoins

- **Collecte des données :** Dans cette étape, nous déterminons les besoins décisionnels initiaux en définissant les informations susceptibles d'intéresser chaque groupe de décideurs. Ainsi on procède à la collecte des requêtes-types pertinentes en interviewant les décideurs, à la mise au point d'un questionnaire permettant de mieux caractériser leurs besoins et à l'analyse des données décisionnelles et des résultats des interviews et des questionnaires afin de dégager les règles de gestion appliquées à ces données.
- **Requêtes-types :** Nous souhaitons spécifier une BDM permettant d'analyser la consommation d'énergie. L'analyse des rapports d'activité existants et une première interview des décideurs a permis de spécifier cinq requêtes-types (R1 à R5). Le tableau présente un exemple de rapports d'activité que nous avons collecté pendant l'analyse.

R1: Analyser	la consommation en fonction des machines, des capteurs et des heures
R2: Analyser	la consommation en fonction des entrepôts, des capteurs et des mois.
R3: Analyser	la consommation en fonction des entreprises, des capteurs et des années.
R4: Analyser	la consommation en fonction des locaux techniques, des capteurs et des mois.
R5: Analyser	les valeurs de mesures en fonction des entreprises, des locaux techniques, des entrepôts, des machines, des capteurs et du temps.

Tableau 1. Requêtes-types

2.1. Spécification des besoins

L'étape de spécification permet d'analyser les données collectées afin de spécifier les besoins des décideurs. Ces besoins seront organisés en termes de paramètres et de mesures afin de préparer la définition du schéma multidimensionnel. Il s'agit de créer la matrice des besoins en fonction de nos requêtes crées auparavant afin d'en tirer les dimensions et les faits.

Parametres /Indicateurs	Id_Machine	Id_Capteur	Heure	Id_Entrepot	Mois	Id_entreprise	Annee	Id_LT
Valeur_cons	x	x	x	x	x	x	x	
Valeur_mes	x	x	x	x	x	x	x	

Tableau 2. Spécification des besoins

2.2. Formalisation des besoins

Après avoir collecté et spécifié les besoins des décideurs, nous réalisons dans cette étape la formalisation de ces besoins sous forme d'un schéma multidimensionnel contraint. Nous commençons tout d'abord, par la définition des faits, puis la définition des dimensions et l'enrichissement, ensuite, nous finissons par la création du schéma multidimensionnel.

2.3. Définition des faits

Cette étape nous permet de regrouper les mesures dans des faits (sujets d'analyse). La définition des faits peut être réalisée d'une manière automatique en regroupant les

mesures analysées au travers de paramètres identiques. Au niveau de cette étape, nous définissons aussi les fonctions d'agrégation compatibles avec chaque mesure.

F1 : Fact_Energie { Val_Energie (sum, avg, max, min), Val_Energie (sum,avg, max, min) }

2.4. Définition des dimensions et enrichissement

Cette étape réalise le regroupement automatique des paramètres dans des dimensions (axes d'analyse), l'enrichissement de ces dimensions à travers de nouvelles propriétés tels que les attributs faibles qui seront greffés aux différentes dimensions de l'analyse et la définition de la granularité de l'analyse en choisissant le niveau d'analyse le plus fin pour chaque dimension.

- **D1 : D_Machine**{ Id_Machine,nom }
- **D2 : D_Capteur**{ Id_Capteur,nom }
- **D3 : D_Temps**{ année,mois,jour,heure }
- **D4 : D_Entrepot**{ Id_Entrepot,nom }
- **D5 : D_Entreprise**{ Id_Entreprise,nom }
- **D6 : D_LocauxTechniques**{ Id_LocauxTechniques,nom }

2.5. Démarche ascendante

Cette démarche part du schéma conceptuel de la source de données pour construire le schéma multidimensionnel contraint d'un magasin de données. Afin de tenir compte des informations contenues dans la source de données, nous proposons une démarche ascendante incrémentale.

Dans cette démarche, nous supposons que le concepteur du schéma multidimensionnel a une double compétence informatique et métier. En effet, c'est le concepteur du schéma multidimensionnel qui doit détecter les différents centres d'intérêt de l'organisation en analysant le schéma de la source.

2.6. Détermination des faits

La détermination des sujets de l'analyse permet de dresser la liste des faits du modèle multidimensionnel. Selon notre démarche, un fait est projeté à partir d'une Classe Représentative (CR) de la source. Une Classe Représentative (CR) enregistre les valeurs numériques et les évènements particuliers.

Pour notre application, la classe la plus active et qui représente un événement date est la classe Mesure. Cette classe représente l'activité des capteurs. A partir de cette dernière, nous définissons le fait (F_mesure) englobant les mesures : "value" permettant de tracer les valeurs générées par les capteurs, "nom_capteur" le nom du capteur,"chrono" le nombre en milliseconde de l'événement depuis 2017, "quality" indique la qualité de la mesure, "type_capteur" précisant s'il s'agit d'un compteur, sonde, vanne motorisée ou débitmètre et la dernière mesure "consommation" c'est une valeur à calculer à partir de deux valeurs de value en fonction du type du capteur.

F1 :Fact_Energie= {value=S.Energie.value (sum, avg, max, min), énergie = (s.Energie.value1-S.energie.value2) }.

Cette classe est choisie par le concepteur de la Base de données Multidimensionnelle. Celui-ci analyse la source et définit les classes susceptibles d'être transformées en sujets d'analyse. Le concepteur choisit comme classes représentatives les classes qui comportent des propriétés numériques transformables en indicateurs d'analyse. Les mesures du fait sont obtenues en appliquant une fonction de calcul sur un ou plusieurs attributs de la classe représentative.

2.7. Détermination des dimensions

Les dimensions sont élaborées à partir des classes sources reliées à la classe représentative, appelées classes déterminantes. La détermination de ces classes est réalisée automatiquement en suivant le principe de dépendance fonctionnelle entre classes. Une classe Ci est déterminante d'une classe CR si et seulement si Ci=Cr ou Cr est relié à Ci par un lien de dépendance fonctionnelle directement ou indirectement. Après une analyse sémantique, nous avons dégagé la dimension "**D_Capteur**" qui regroupe les paramètres d'analyse et la dimension "D_temps" présentées ci-dessous:

ID_Capteur{
id_capteur=S.Capteur.id_capteur,
nom_capteur = S.Capteur.nom_capteur,
unité = S.Capteur.unite_capteur,

id_Entreprise = S.Entreprise.id_ Entreprise,
nom_ entreprise = S.Entreprise.nom_ entreprise,
reference =S.Entreprise.reference_ entreprise,
adresse =S.Entreprise.adr_ entreprise,

CodePostal = S.Entreprise.code_postal,
ville = S.Entreprise.ville,
localisation = S. Entreprise.localisation,
responsable =S.Entreprise.responsable,
actif = S.Entreprise.Actif,

id_Entrepôt = S.Entrepot.id_ entrepôt,
nom_ b = S.Entrepot.nom_ entrepot,
ref_b = S.Entrepot.Ref_ entrepot,
type = S.Entrepot.type_ entrepot,
surface = S.Entrepot.surface_ entrepot,
nbr_niveau =S.Entrepot.nbr_niveau,
gerant = S.Entrepot.gerant,

id_LT = S.LocauxMachines.id_local,
nom_local =S.LocauxMachines.nom_local,
ref_local = S.LocauxMachines.ref_local,
réseau = S.LocauxMachines.reseau_local,
type = S.LocauxTechnique.type,

id_machine = S.Machine.id_machine,
nom_machine =S.Machine.nom_ machine,
marque =S.Machine.marque_ machine,
Nsérie = S.Machine.Nsérie,}
Puissance= S.Machine.puissance_machine,

D_Temps{Timestamp=S.Capteur.TS}

2.8. Définition de la dimension temporelle

La dimension temps est basée sur l'attribut source TS stockant la date de chaque mesure du capteur et qui a le format 'DateTime'. Dans notre contexte d'entrepôt de données, nous n'avons pas besoin de créer une dimension temps destinée à historiser un enregistrement pour chaque minute ou chaque seconde d'une période entière. Nous avons donc besoin de la granularité heure, c'est pourquoi nous l'avons définie comme la granularité d'analyse.

DTemps: {
codeDate = S. Energie.codeDate,
Heure = S. Energie.Heure,
Jour= S. Energie.jour,
Mois = S. Energie.Mois,
Année = S. Energie.annee }

4.3 Démarche mixte

La demarche mixte est basée sur deux étapes :

-la première étape consiste à construire un schéma multidimensionnel qui comporte les données communes: fait, dimension, hiérarchies, ... ;

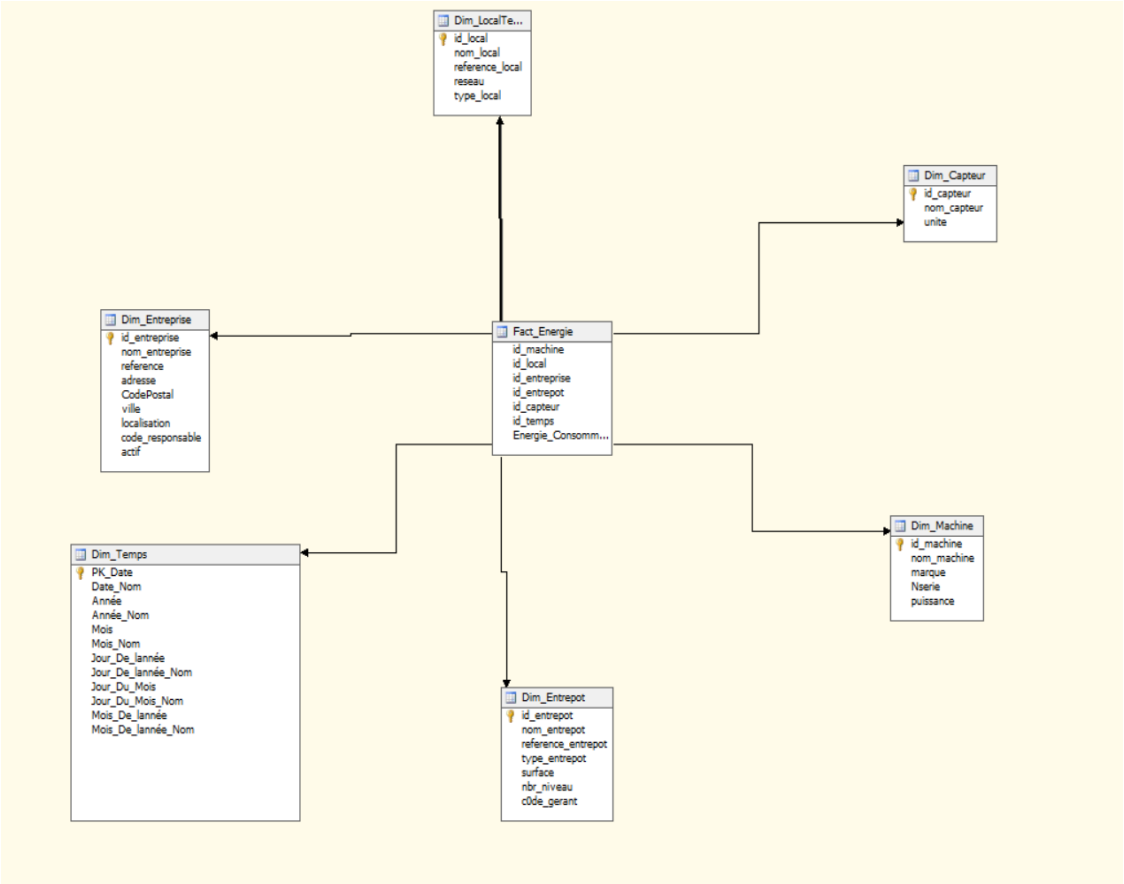

Figure III.4 : Schéma multidimensionnel

3. Environnement et Outils

3.1. Environnement Technique

Dans cette partie nous posons d'abord les capteurs sur les machines des l'entreprises du Domaine Industriel de Diamniadio en respectant les critères suivants :

- **Typologie des capteurs** : Les Capteurs plug & play posés, qui ne demande aucun paramétrage à effectuer. Ils remontent les données relevées directement sur le logiciel disponible sur smartphone et ordinateur.
- **Typologie des énergies :** La diversité des capteurs connectés vous permet de suivre toutes les énergies : électricité, gaz, eau, température, qualité d'air, humidité…
- **Large couverture réseau** : Les capteurs capable de communiquer avec le **réseau basse** fréquence d'Orange Telecom, qui couvre 95% de la population
- **Haute précision :** Les capteurs ont une précision inférieure à 0,2°C pour les températures, et inférieure à 2% pour les capteurs d'énergie afin de recevoir des données fiables et facilement exploitables.

3.1.1. Architecture d'un capteur sans fil

Un capteur est composé de quatre unités principales, qui sont présentées dans la figure.

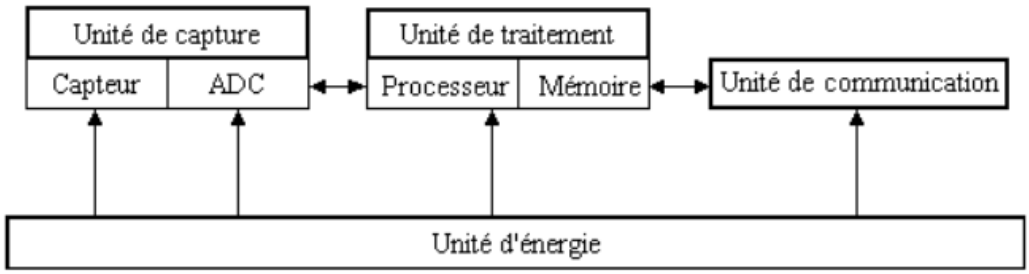

figure 11. Architecture d'un capteur

- **Unité de capture (Sensing unit) :** elle est composée de deux sous unités, un dispositif de capture physique qui prélève l'information de l'environnement local et un convertisseur analogique/numérique appelé ADC (Analog to Digital Converters).
- **Unité de traitement (Processing unit):** les données captées sont communiquées au processeur où elles sont stockées dans la mémoire.
- **Unité de communication (Transceiver unit):** elle est composée d'un émetteur/récepteur (module radio) permettant la communication entre les différents nœuds du réseau.

- **Unité d'énergie (Power unit):** c'est la batterie qui, n'est généralement ni rechargeable ni remplaçable. La capacité d'énergie limitée au niveau des capteurs représente la contrainte principale lors de la conception de protocoles pour les réseaux de capteurs.

Les unités d'énergie peuvent être supportées par des photopiles qui permettent de convertir l'énergie lumineuse en courant électrique.

3.1.2. Architecture de communication

Les capteurs sont habituellement dispersés dans une zone de capture. Chacun de ces capteurs a la possibilité de collectées les données et de les router vers une ou plusieurs stations de base. Ce dernier est un point de collecte de données capturées. Il peut communiquer les données collectées à l'utilisateur final à travers un réseau de communication, éventuellement l'Internet. L'utilisateur peut à son tour utiliser la station de base comme passerelle, afin de transmettre ses requêtes au réseau. Cette architecture est illustrée dans la figure IV1.2.

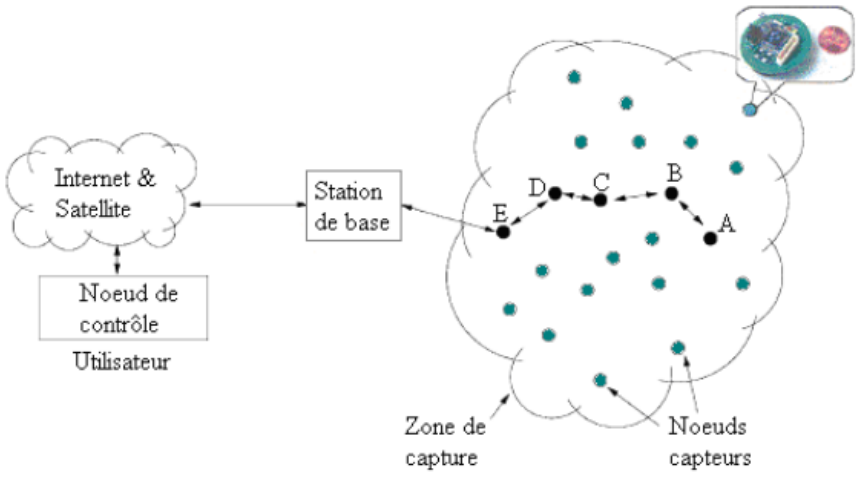

figure 12. Architecture de communication

3.1.3. Applications des Réseaux capteurs sans fil

Les réseaux de capteurs sans fil ont été classés parmi les 21 technologies les plus importantes du 21éme siècle [WEB01]. En effet, la recherche dans le domaine des capteurs est en train de vivre une révolution importante, ouvrant des perspectives d'impacts significatifs dans de nombreux domaines.

Ainsi, nous classifions les applications en quatre classes d'applications : orientées temps (time driven), orientées événements (event driven), orientées requêtes (query driven) et hybride :

Applications orientées temps : Cette classe représente les applications où l'acquisition et la transmission des données capturées sont liées au temps : instant précis, période d'acquisition. Cette période d'acquisition peut être plus au moins longue selon l'application (de quelques secondes jusqu'à quelques heures voire des jours). Ainsi, la quantité de données échangée dans le réseau dépend de la périodicité des mesures à effectuer sur l'environnement local.

Applications orientées événements : Dans ce cas, les capteurs envoient leurs données seulement si un événement spécifique se produit. On peut citer l'exemple de surveillance des alarmes dans une entreprise où un capteur envoi des signaux à la station de base. Au départ, cette classe d'application était conçue à des fins militaires, comme la surveillance du déplacement d'objets dans le champ de bataille. Par la suite, cette classe a rapidement trouvé de nouvelles perspectives comme le contrôle industriel, le contrôle médical des patients, la surveillance d'édifices (barrages, ponts, voies de chemins de fer, etc.)

Applications orientées requêtes : Dans ce cas, un capteur envoi de l'information uniquement suite à une demande explicite de la station de base. Cette classe d'application est destinée aux applications adaptées à l'utilisateur. Ce dernier peut requérir des informations à partir de certaines zones dans le réseau ou interroger les capteurs pour acquérir des mesures d'intérêts. Dans ce cas, des connaissances sur la topologie du réseau et l'emplacement des capteurs sont nécessaires.

Applications hybrides : Ce type d'application met en œuvre les trois modes de fonctionnement décrits précédemment. Par exemple, dans un réseau conçu pour le suivi d'objets, le réseau peut combiner entre un réseau de surveillance (time driven) et un réseau de collecte de données par événements (event driven). Par exemple, pendant les longues périodes d'inactivité des capteurs et lorsque aucun objet n'est présent, le réseau peut assurer une fonction de surveillance.

3.1.4. Consommation d'énergie dans les Réseaux capteur sans fil

L'énergie consommée par un capteur est due essentiellement aux opérations suivantes : la capture, le traitement et la communication de données.

Energie de capture : L'énergie de capture est dissipée pour accomplir les tâches suivantes : échantillonnage, traitement de signal, conversion analogique/numérique et activation de la sonde du capture. En général, l'énergie de capture représente un faible pourcentage de l'énergie totale consommé par un capteur.

Energie de traitement : L'énergie de traitement se divise en deux parties : l'énergie de commutation et l'énergie de fuite. L'énergie de commutation est déterminée par la tension d'alimentation et la capacité totale commutée au niveau logiciel (en exécutant un logiciel). Par contre l'énergie de fuite correspond à l'énergie consommée lorsque l'unité de calcul n'effectue aucun traitement. En général, l'énergie de traitement est faible par rapport à celle nécessaire pour la communication.

Energie de communication : L'énergie de communication se décline en deux parties : l'énergie de réception et l'énergie de l'émission. Cette énergie est déterminée par la quantité des données à communiquer et la distance de transmission, ainsi que par les propriétés physiques du module radio. L'émission d'un signal est caractérisée par sa puissance. Quand la puissance d'émission est élevée, le signal aura une grande portée et l'énergie consommée sera plus élevée. Notons que l'énergie de communication représente la portion la plus grande de l'énergie consommée par un nœud capteur.

3.2. Environnement Logiciel

Plusieurs solutions ETL open source existent sur le marché nous avons mené une étude sur quelques solutions open source pour choisir l'outil qui convient à notre besoin. Nous avons choisi de faire une comparaison entre les deux outils les plus utilisés à savoir ; SQL Server Management Studio, Pentaho [16] Data Intégration (PDI), Oracle et Talend open Studio (TOS).

Dans ce chapitre nous allons présenter les outils les plus utilisés suivants leurs utilités, fonctionnalités et avantages. Tous ces outils sont fiables, performants et conviviaux mais étant donné que SQL Server management studio est une plateforme de gestion des données plus générale et complète, nous avons opté de l'utiliser dans la partie intégration. Dans ce qui suit nous décrivons la suite d'outil utilisés dans notre architecture.

3.2.1. Pentaho Data Integration

Pentaho [16] est un outil d'entrepôts de données utilisé pour la migration de données d'une base à une autre, l'alimentation d'un Data Warehouse et de Data Marts [17]. Pentaho [16] est une solution d'informatique décisionnelle op en source entièrement développée en Java. Elle porte sur toute la chaîne décisionnelle et utilise différents outils et composants:

- Pour la collecte et l'intégration: les outils d'ETL(Extract-Transform-Load) Kettle ou Mon-darian
- Pour la diffusion: un serveur d'application JBossou TOMCAT.
- Pour la présentation: JFreeReport, BIRT ou encore Jasper Report.

3.2.2. Talend Open Studio

Talend open studio est un outil d'entrepôts de données utilisé pour l'échange Inter-Application (EAI), la migration de vieux applicatifs vers de nouveaux, le contrôle de l'information et l'alimentation du décisionnel.

Talend propose la seule plateforme unifiée simplifiant la gestion de données et l'intégration d'applications en fournissant un environnement unifié pour gérer le cycle de vie complet à travers les frontières de l'entreprise.

3.2.3. SQL Server management studio

Apres la présentation des différents outils d'implémentation les plus utilisées d'un data warehouse à savoir SQL Server, Oracle, Pentaho [16] et Talend Open Studio, en spécifiant: leurs définitions et fonctionnalités principales. Dans la partie suivante nous porteront notre choix sur SQL Server qui est l'outil le plus adapté pour notre application.

SQL Studio offre une multitude d'outils puissants pour l'administration de bases de données, la gestion des objets des bases de données, pour la migration, la comparaison, l'extraction des bases de données, pour l'importation, l'exportation et la comparaison des données.

3.2.4. SQL Server Integration Services SSIS

Afin de garder un système d'information cohérent et de simplifier la vie des informaticiens, Microsoft SQL Server 2008 est doté d'un nouvel outil d'intégration de données (d'une source en passant par une transformation optionnelle des données, vers

une destination) appellé **SQL Server Integration Services (SSIS).** Il permet de communiquer des données de systèmes d'information différents ce qui représente bien les besoins actuels en terme de système d'information, de transfert de données et permet de garder une cohérence.

SSIS est un outil d'extraction, de transformation et de chargement de données, en bref ce que l'on appelle un ETL. On extrait d'une source de données, puis suit la transformation si besoin, pour ensuite injecter ces données vers MS SQL Server ou encore d'autres destinations.

L'environnement de conception d'un package SSIS est donc Visual Studio avec, si possible, un accès à votre serveur de données. Ceci en vue au minimum de contrôler que l'import s'est bien effectué (en plus des logs de progression disponibles avec votre package SSIS).

3.2.5. NetBeans

NetBeans est un environnement de développement intégré, ou IDE, pour la création de programmes d'ordinateur dans un certain nombre de langues différentes. Le développement NetBeans se réfère au processus de l'utilisation de NetBeans pour créer, éditer et organiser le code qu'on développe dans un programme informatique.

NetBeans soutient le développement en Java, PHP, HTML, JavaScript, CSS et C + +. Dans ce contexte, le développement se réfère à un codage à l'appui, le débogage et la compilation du code dans ces langues. Dans la phase de codage, NetBeans vérifie le code en temps réel pour assurer une syntaxe correcte. NetBeans comprend à un certain nombre d'outilsde débogage pour nous aider à isoler et corriger les bugs non liées à la syntaxe. NetBeans peut aussi compiler le code dans un programme de travail. Pour que notre code soit compilé il faut que le logiciel soit installé avec le kit de développement associé à chaque version. Le Java Développement Kit, communément appelé JDK, est le kit de développement de base que propose gratuitement la société Oracle.

3.2.6. Power AMC

PowerAMC est la version française de PowerDesigner. C'est un outil de conception informatique qui peut élaborer plusieurs modèles de données et rétro générer des codes dans divers langages de programmation. Supportant la conception suivant la méthode Merise, il peut mettre en œuvre des prototypes de données utiles à la conception de diverses applications. Il supporte le travail en groupe par le biais d'un outil référentiel, offrant l'opportunité de régir le travail d'équipe sur une plate-forme client/serveur. Connu à ses débuts sous le nom AMCDesignor, il servait d'outil de modélisation graphique pour Oracle. Il a ensuite été racheté

par Powersoft qui l'a traduit en anglais avec PowerDesigner et en a adapté la méthodologie IE (Information Engineering).

4. Réalisation

4.1. Etape de la realisation

Pour la création du Data Warehouse, nous avons suivi les trois étapes majeures montrées dans la figure ci-dessous, commençant par la création de modèle dimensionnel, ensuite la création de cube OLAP [18] et a la fin la représentation graphique.

- pour la création de modèle dimensionnel, nous avons commencé par la création de la base de données de l'entrepôt, ensuite la création et la peuplassions de la table de fait et les tables de dimensions.
- pour la création de cube OLAP [18], nous avons établi la source des données et des Vues, crée le cube, puis l'attribut hiérarchique dans les dimensions, déployer le cube pour un éventuel traitement.
- pour la représentation graphique, nous avons effectué une connexion entre notre Cube de données et l'outil graphique Excel.

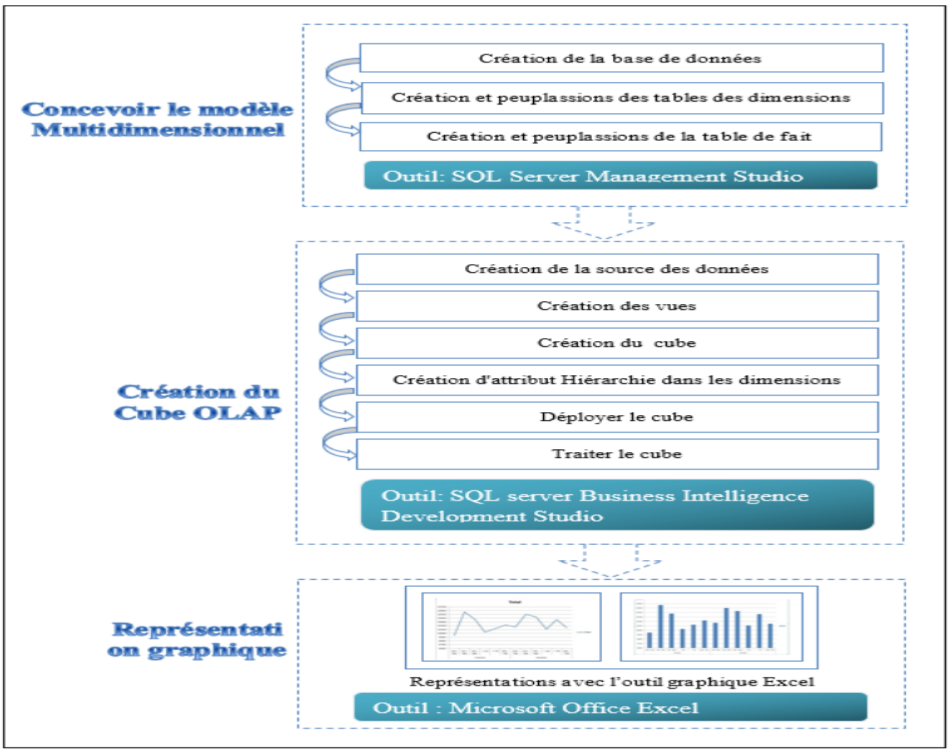

figure 13. Les étapes de la conception

4.2. Création Le modèle multidimensionnel avec SQL Server Management Studio

Pour la création de la base de données pour notre entre Data Warehouse, nous avons utilisé l'instruction présente en ANNEXE 1. Dans ce schéma, on voit bien qu'on a la table de fait centrale appelée **FactEnergie** contenant toutes les mesures qu'on veut analyser. Et autour on voit toutes les tables de dimension. Ce schéma est appelé schéma en étoile parce que les tables de dimension sont toutes reliées directement à la table de fait.

Notre base de données ETUDE a une structure multidimensionnelle mais reste cependant une base de données relationnelle. Car elle est gérée par le moteur de base de données relationnel de Sql Server (Relational DataBase Engine). Par contre nous créerons aussi une structure multidimensionnelle (qui sera notre cube Analysis Services) sous le serveur d'analyse de Sql Server et là c'est le moteur OLAP [18] qui intervient.

Ce moteur se chargera donc de faire les pré-aggrégations et aggrégations nécessaires. Cela nous amènera donc à faire 2 travaux:

- Un premier travail d'optimisation de la base ETUDE (en choisissant la bonne modélisation et les bon indexes)
- Un deuxième travail d'optimisation du cube de données (en choisissant la bonne modélisation, les bonnes représentations des hiérarchies utilisateurs et les bonnes relations d'attributs etc.)

Maintenant nous allons donc entrer sous Analysis Services et construire notre cube de données ou base de données multidimensionnelle.

4.3. Création de cube OLAP avec SQL Business Intelligence Management studio

Pour le démarrage de projet analyses services nous avons cliqué sur "fichier" ensuite" nouveau" et "projet" après nous avons sélectionné "projet analyses services" et

à la fin nous avons donné un nom pour le projet qui s'appelle « ETUDE » (Figure IV.3.1)

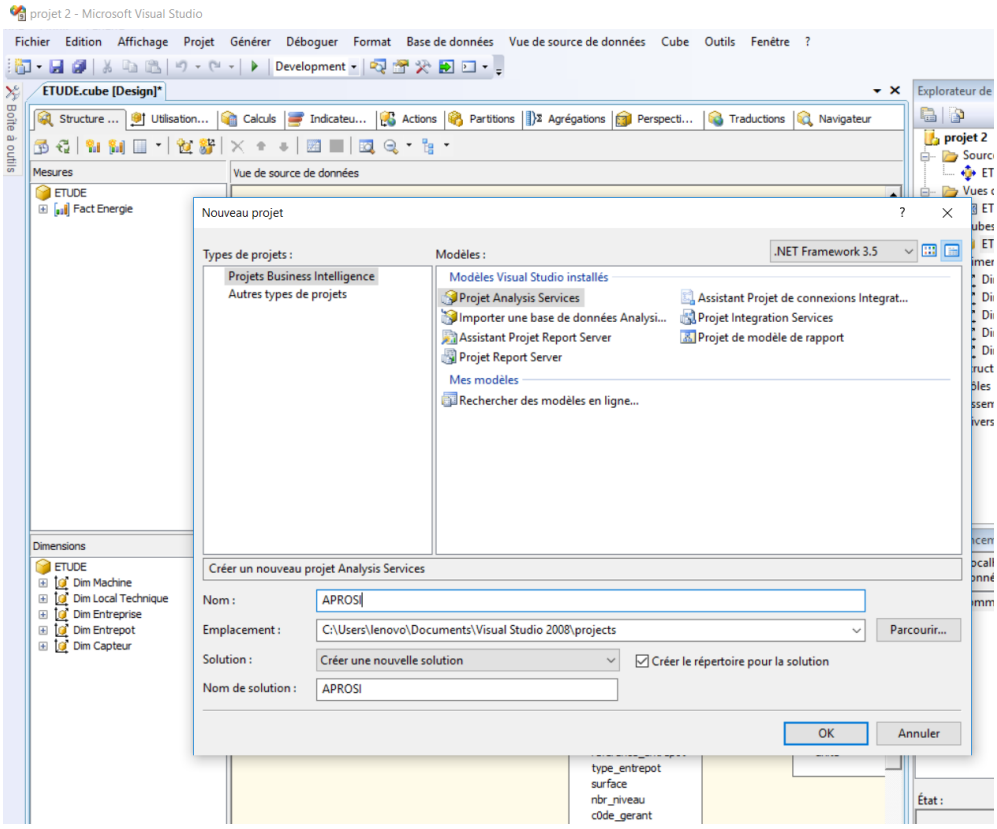

figure 14. Démarrage de Projet Analyses services

4.4. Création de la source de données

Pour la création de la source de données, nous avons suivi les étapes suivantes:

- Nous avons cliqué sur le bouton droit sur "source de données" dans l'explorateur de solutions situé au côté droit de l'interface, la (Figure IV.3.2) montre l'interface. La première chose à faire c'est de créer une source de données sur notre entrepôt de données ETUDE Pour cela, dans la boîte de dialogue Connection Manager, choisir le fournisseur de données Native OLE DB\SQL Native Client, le nom du serveur contenant votre base de données ETUDE (dans mon cas, je l'ai mis sur mon ordinateur en local), le type d'authentification (moi j'ai choisi le mode authentification windows) et enfin, choisissez la base de données ETUDE.

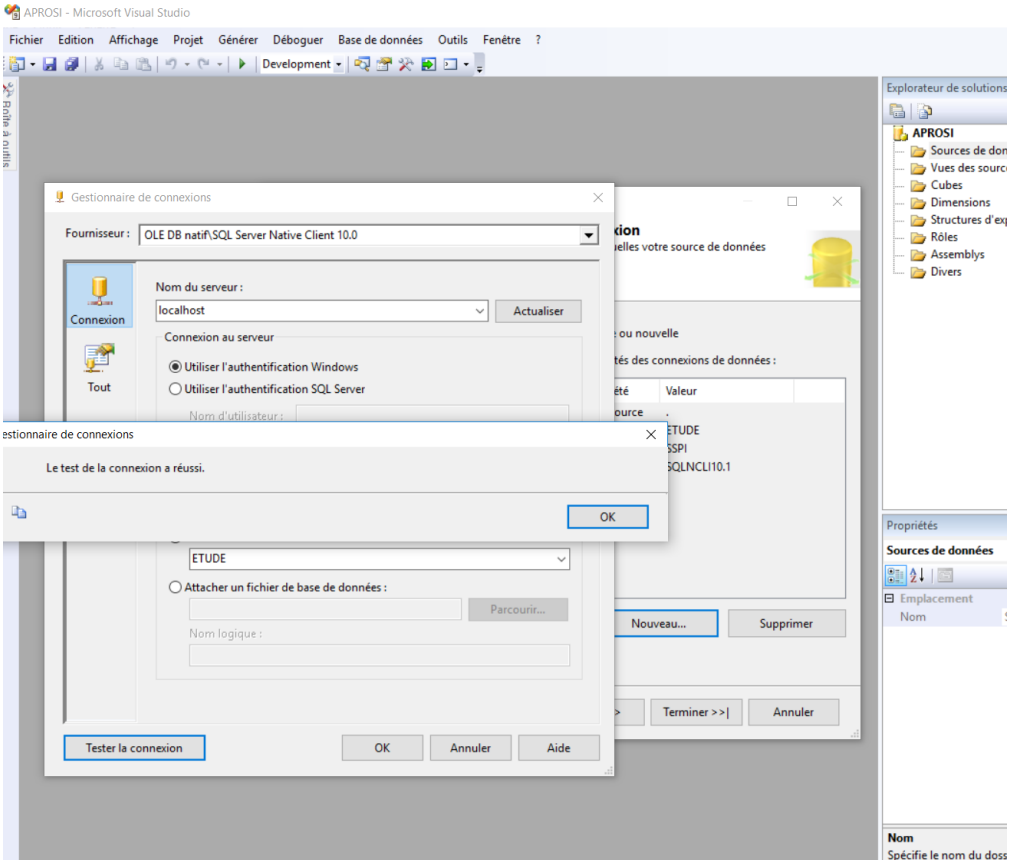

figure 15. Nouvelle source de données

Il faut aussi choisir le login utilisé par Analysis Services pour opérer sur la base ETUDE qui s'agir du compte système local.

4.5. Création de la vue de source de données

La deuxième chose à faire est de créer une vue de source de données sur notre source de données. La vue de source de données sert à sélectionner des tables de la source de données. Ainsi pour une même source de données, on peut avoir plusieurs vues de sources de données (une par type de besoin par exemple). Nous allons donc créer une source de données qui contiendra les tables qui seront nécessaires à notre analyse.

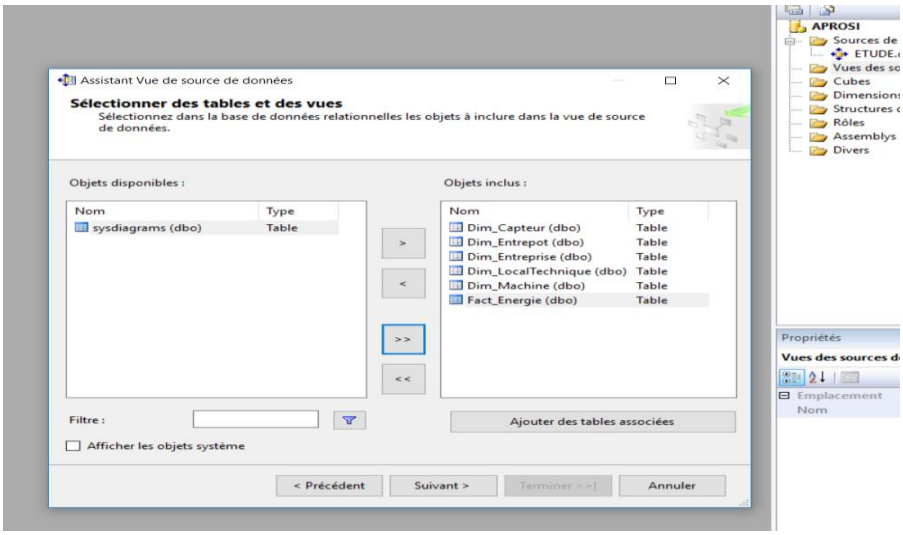

figure 16. Schéma sélection des tables de vues

Une fois avoir ajouté les tables dont nous avons besoin, on obtient le schéma suivant. Les liens ont été créés par Analysis Services. Ce dernier se base sur les clés étrangères définies au niveau de la base de données ETUDE pour déterminer les liens dans Analysis Services.

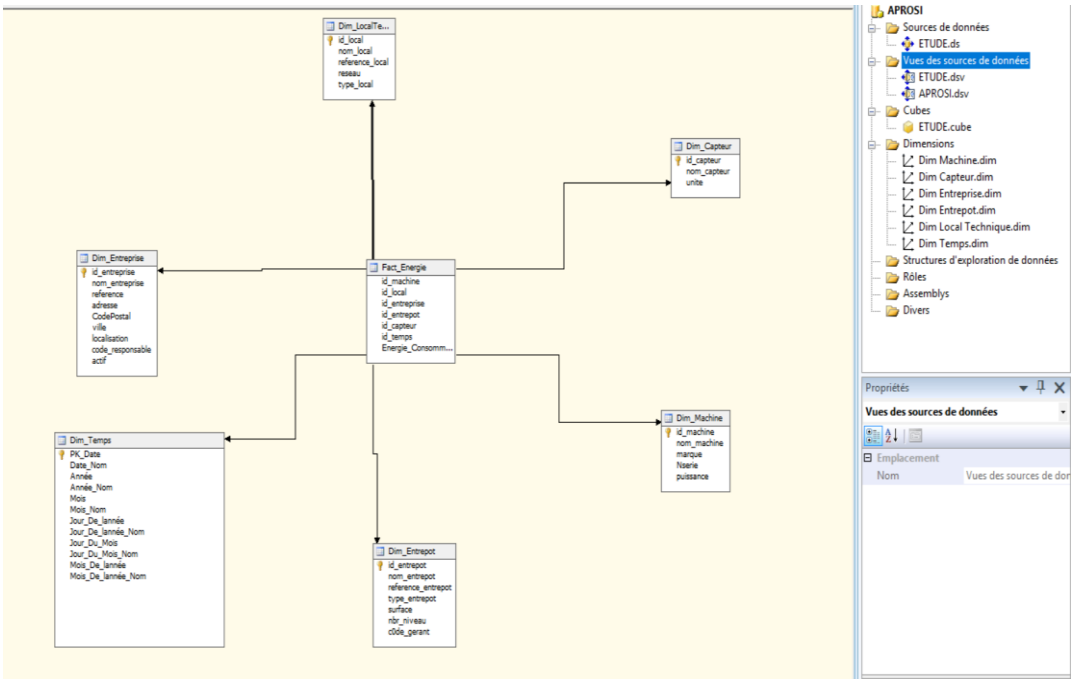

figure 17. Vues Source de données

4.6. Création du cube

Maintenant que nous avons la vue sur les données que nous allons utiliser, nous allons créer notre cube de données. Ici nous allons créer les dimensions proprement puis, traitez les, créez les mesures et traitez votre cube.

Comme le montre la figure suivante, le cube sera construit en utilisant notre source de données définie. De plus, Analysis Services construira automatiquement ses attributs et ses hiérarchies en s'appuyant sur les métadonnées de la base ETUDE. Bien évidemment, pour avoir un cube optimal il ne faut pas laisser Analysis Services s'en occuper.

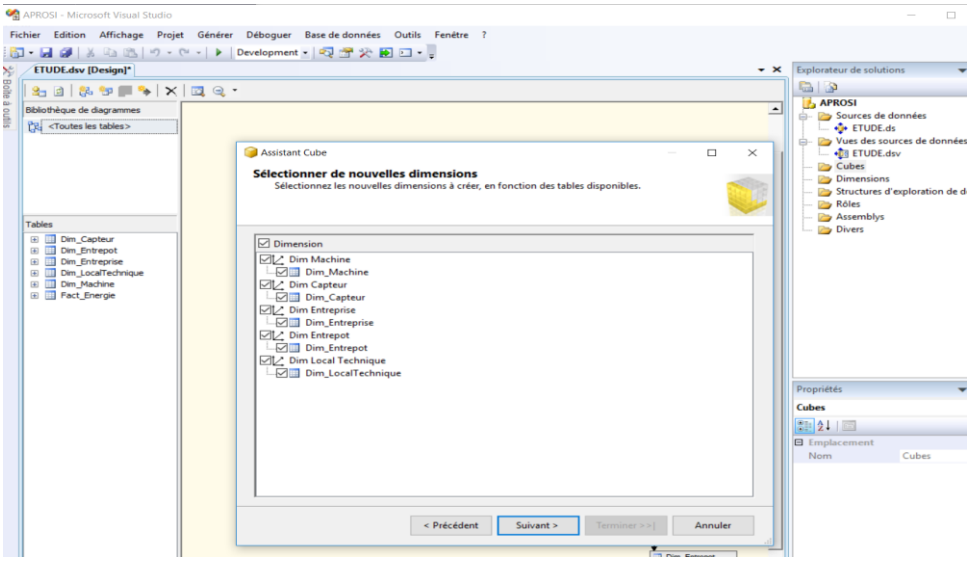

figure 18. sélection des nouvelles dimensions

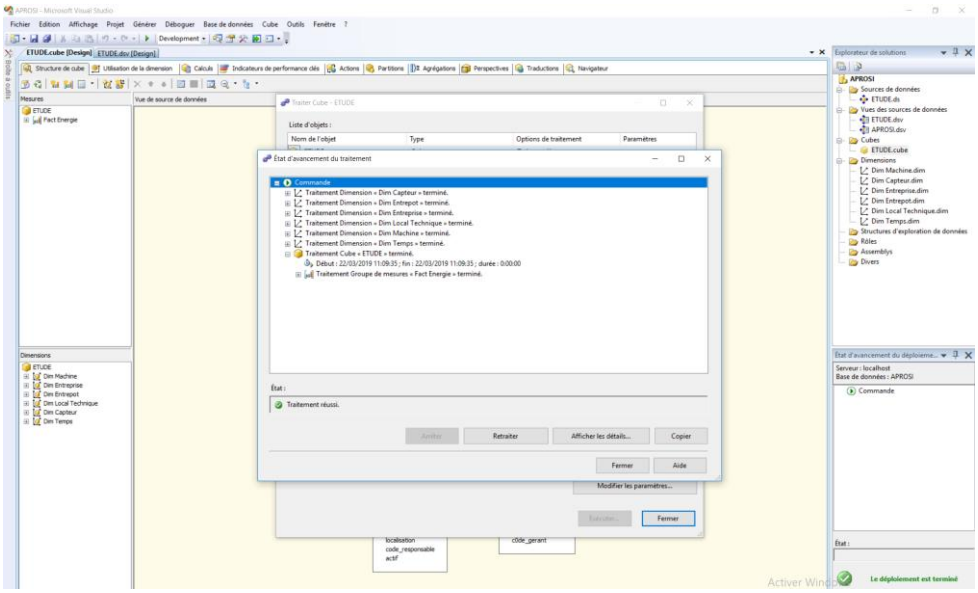

figure 19. Nouveau Cube

5. Déploiement

Le déploiement est le point de convergence de la technologie, des données et des applications utilisateur accessibles à partir du poste de travail. Une planification de grande envergure est indispensable afin que toutes les pièces du puzzle se placent correctement. Il est également nécessaire de prévoir une formation des utilisateurs qui intègre tous les aspects de cette convergence. En outre, avant de permettre à l'utilisateur d'accéder à l'entrepôt de données, il faut mettre en place les processus de communication, de support utilisateur, et de prise en compte des demandes d'évolution ou de correction.

5.1. Déploiement de Cube

Le Déploiement de cube se fait en 3 étapes suivantes:

Etape 1:Dans l'explorateur des solutions, nous avons effectué un clique droit sur le nom du projet "ETUDE", ensuite sélectionné" Propriétés".

Etape 2: Dans "Propriétés de configuration" nous avons sélectionné "Déploiement".

Etape 3: Une fois que le déploiement est terminé, nous pouvons voir un message "terminé" dans les propriétés de déploiement.

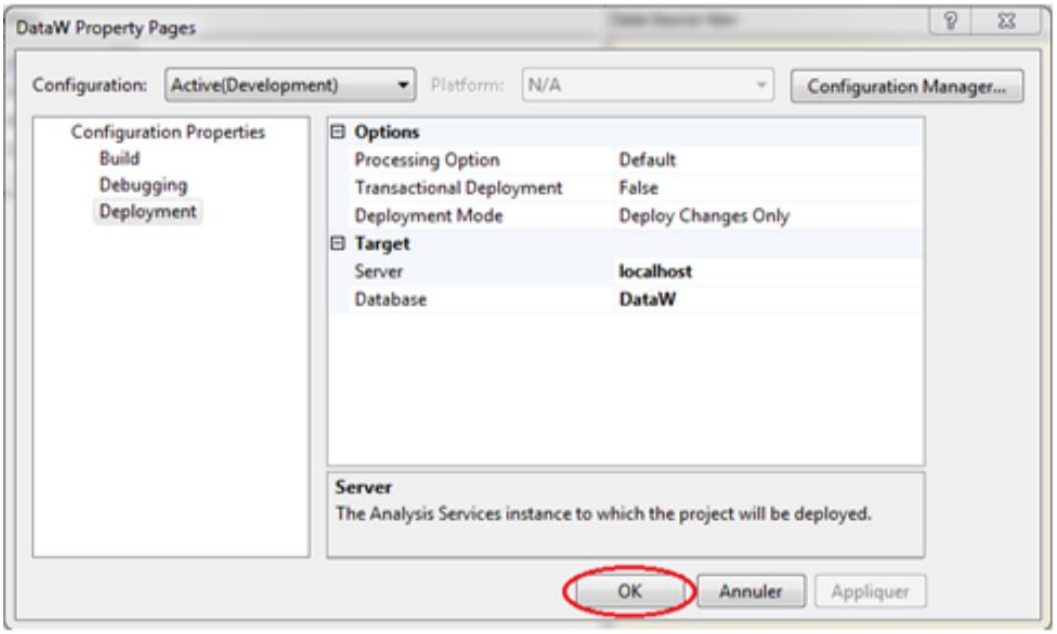

figure 20. Déploiement du Cube

5.2. Traitement de cube

Le Traitement de cube ce fait en 3 étapes suivantes:

Etape 1: Dans l'Explorateur de solutions, nous avons effectué une clique droite sur le nom du projet "ETUDE" ensuite une clique sur "processus".

Etape 2: Exécuter pour traiter le Cube

Etape 3: Une fois que le déploiement est terminé, nous pouvons voir un message "terminé" dans les propriétés de déploiement.

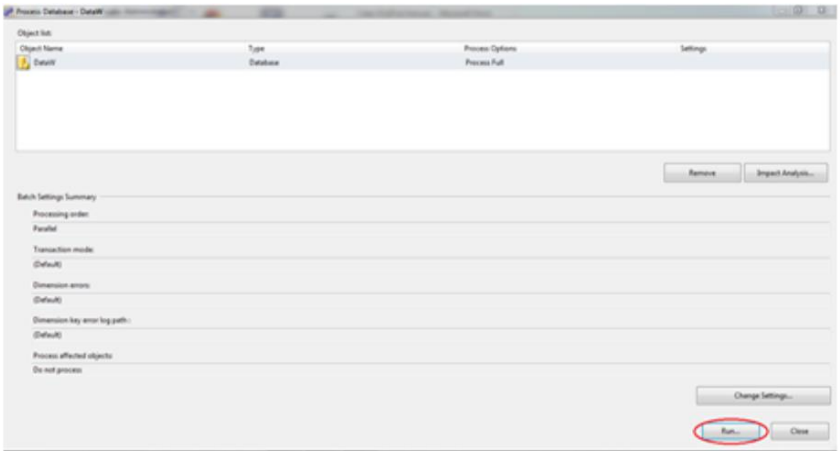

figure 21. Exécution

Etape 3: Une fois le traitement est terminé, nous pouvons voir l'état comme "Successeur" Processus .

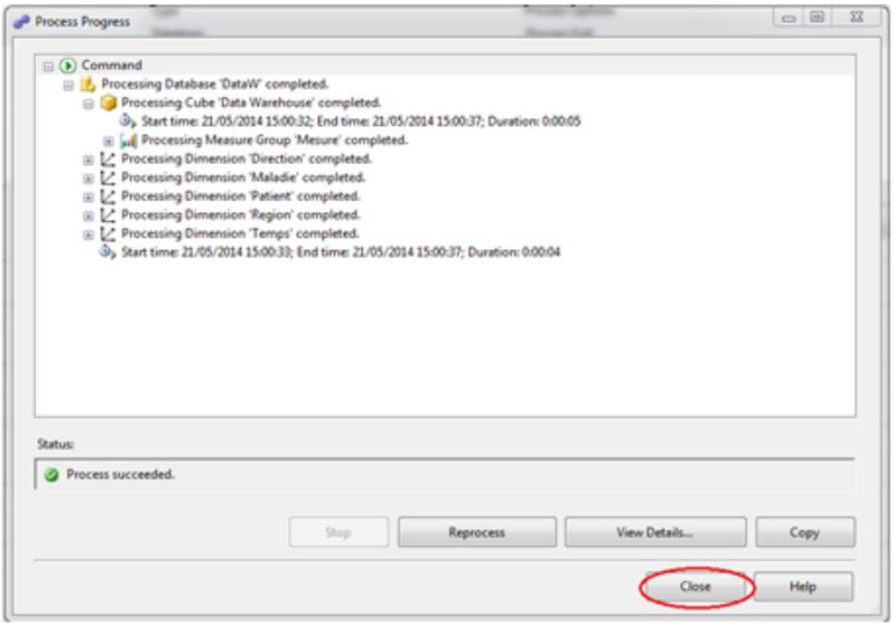

figure 22. Fin de Traitement

5.3. Exploration du Cube Navigateur

L'explorateur Navigateur du Concepteur de cube permet d'explorer les dimensions, les mesures et les indicateurs de performance clés (KPI) d'un cube. Dans SQL Server 2008, le navigateur de cube Analysis Services est intégré au concepteur de requêtes MDX (Multi Dimensional eXpression) et fournit des interfaces utilisateur graphiques pour vous aider à créer des requêtes MDX, filtrer et découper les cubes et descendre dans les hiérarchies.

Le Navigateur comporte deux modes : le mode Création et le mode Requête.

Dans chaque mode, vous pouvez utiliser les objets du volet Métadonnées pour explorer le cube, ou nous pouvons faire glisser des membres du volet Métadonnées vers une zone de requête pour créer une requête MDX qui extraie les données que vous voulez utiliser. L'illustration suivante montre des interfaces du Navigateur en mode Création graphique

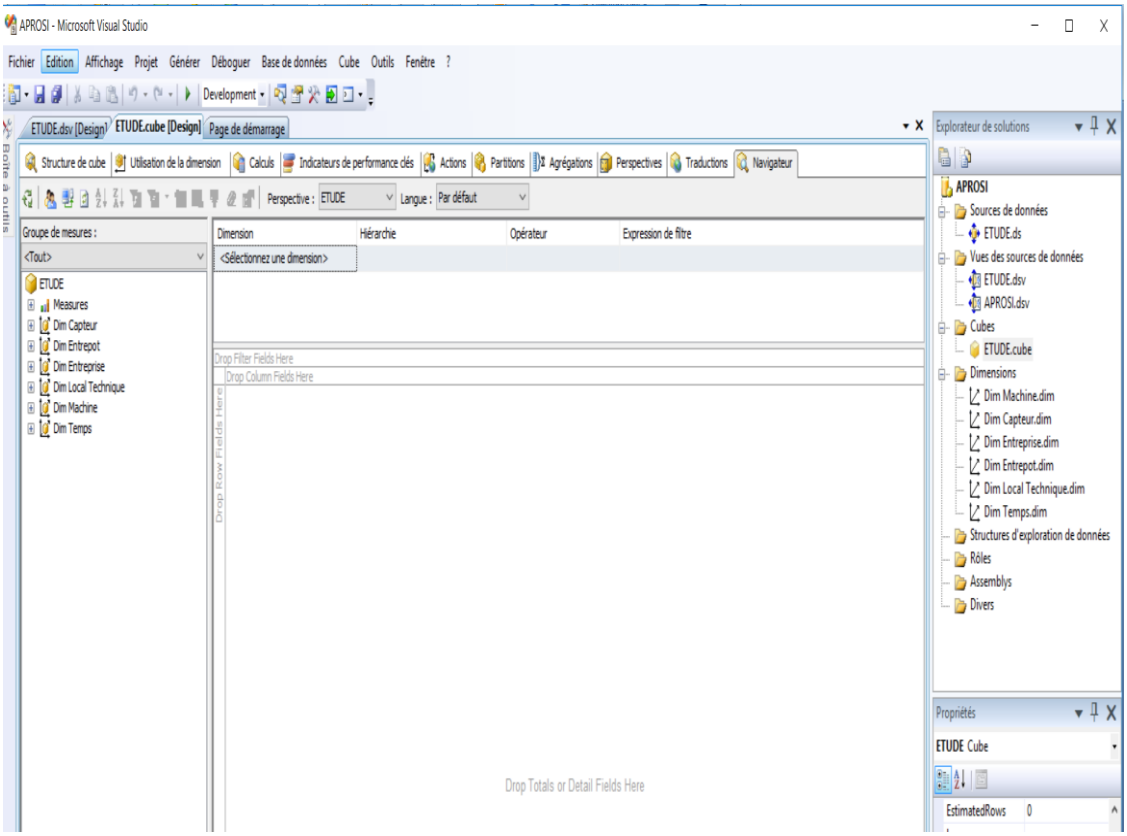

figure 23. Exploration du Cube Navigateur

6. Représentation graphique avec Microsoft Office Excel

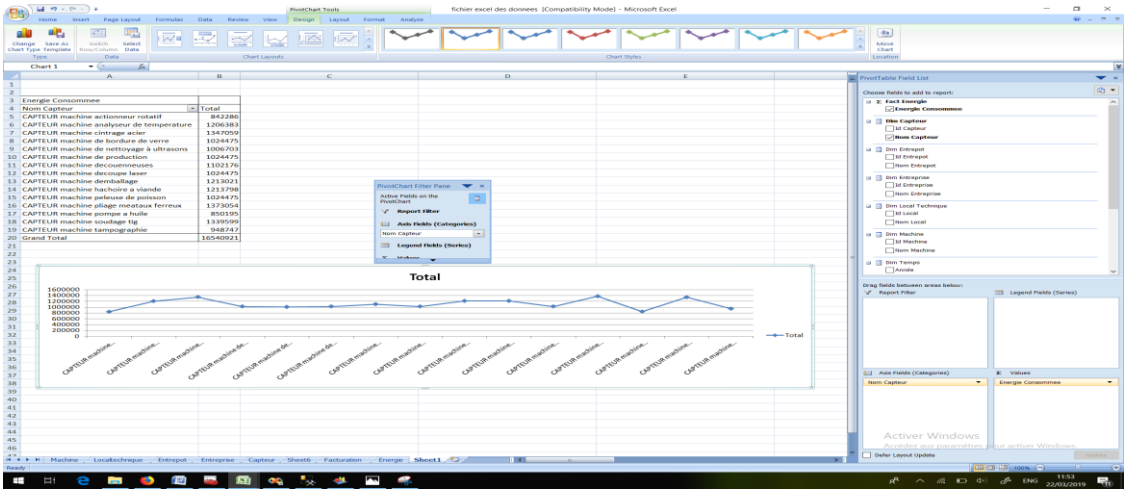

figure 24. Consommation d'énergie par Capteur

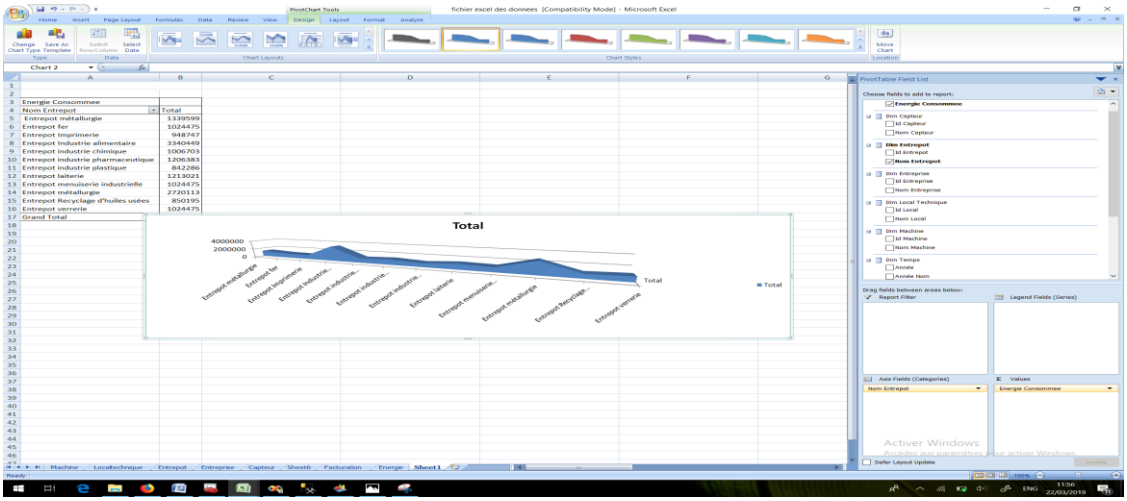

figure 25. Consommation d'énergie par Entrepôt

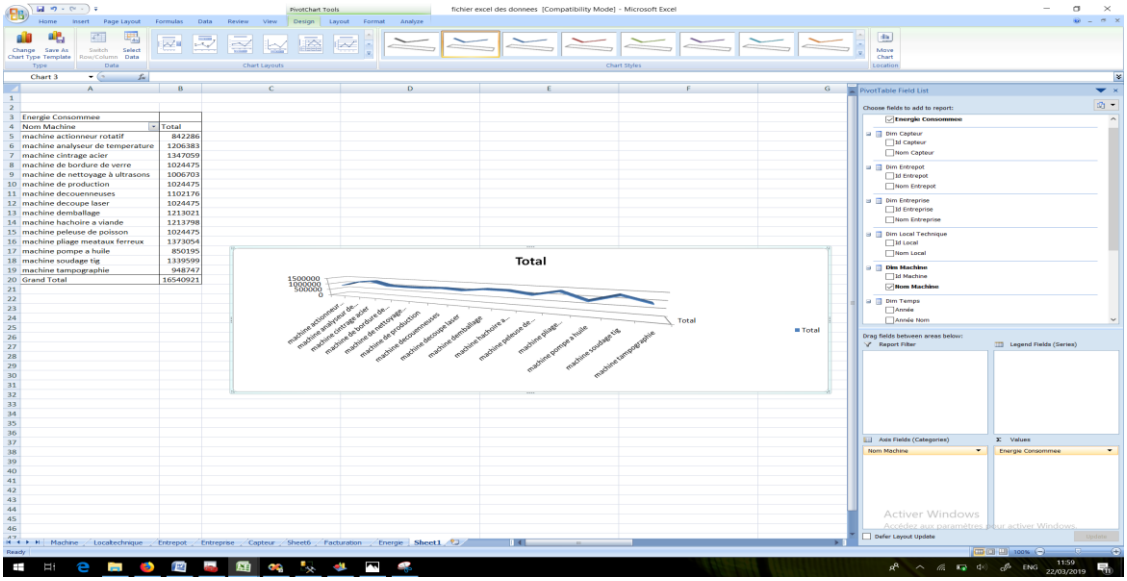

figure 26. Consommation d'énergie par Machine

7. Cout du projet

Un projet est une action spécifique, nouvelle, qui structure méthodiquement et progressivement une réalité à venir, pour laquelle on n'a pas encore d'équivalent.

Le cas de notre projet est un ensemble d'actions à réaliser pour satisfaire un objectif défini, a savoir la conception d'un Data warehouse décision dans le cadre d'une mission précise, et pour la réalisation desquelles on a identifié non seulement un début, mais aussi une fin conforme à des exigences spécifiques telles que des contraintes de délais, de coûts et de ressources.

Les Phases	**Couts Materiels**	**Couts Humain**	**Couts Finaciers**	**Total Couts**
Identifier les besoins auprès des utilisateurs	16 Machines de l'entreprise	2 Charges du projet	500 000 x 2 x 5 mois	5 000.000
Mise en Oeuvre	-16 capteurs dont 1 capteur par machine -1 ordinateur Serveur	5 Ingenieurs	-515.890Franc par capteur -500.000 l'ordinateur - 500 000 x 5 x 5 mois - Charge divers : 3 000 000	16 515.890
Choisir les outils logiciels	1 Logiciel sql server 1 logiciel NetBeans	x	Open Source Open source	
				21 515 000 FCFA

figure 27. Coût du projet

CONCLUSION GENERALE

Le travail qui nous a été confié est la conception d'un Data Warehouse en utilisant une demarche structurée dans le choix des outils d'implémentation, ensuite la conception du système et en fin le déploiement.

Ce travail consiste notre première expérience dans la mise en œuvre d'un Data Warehouse [14]. Nous avons tout de même rencontré beaucoup de difficultés qui nous ont empêchés, en quelque sorte, d'atteindre des résultats satisfaisants à savoir: le manque de documentations, le manque de temps et l'absence des informations au niveau du Domaine Industriel de diamniadio.

En effet dans le déroulement de notre étude, nous avons acquis une bonne expérience dans le domaine des SID basés sur le DataWarehousing. Nous avons aussi bénéficié des connaissances concernant certains outils d'implémentation d'un Data Warehouse (Oracle, Pentaho [16] et Talend Opens Studio) et pour le SQL Server des connaissances plus détaillées.

Dans le futur on aimerait avoir un système d'information opérationnelle (SIO) au niveau du Domaine Industriel de Diamniadio , pour pouvoir construire un système d'information décisionnel (SID) avec des données réels, et aussi, améliorer la sécurité et utiliser des fonctionnalités plus performante pour les Zones Economiques spéciales (ZES) administrées par l'APIX du Sénégal.

BIBLIOGRAPHIE :

1. R. Kimball, L. Reeves, M. Ross, W. Thornthwaite "Le data warehouse: Guide de conduite de projet" axé sur la question de la modélisation Broché – 23 février 2005
2. Emmanuel Ferragu (Auteur)Modélisation des Systèmes d'Information Décisionnels : Techniques de modélisation conceptuelle et relationnelle des entrepôts de données Broché – 27 septembre 2013
3. David maisons, "DATAWAREHOUSE et DATAMINING", , CENTRE DE VERSAILLES, Edition, 11 décembre2006.
4. R. Kimball Concevoir et déployer un Data Warehouse . Guide de conduite de projet Broché – 17 octobre 2000
5. Nacer Boudjlida, « Gestion et administration des bases de données », Application à Sybase et Oracle, DUNOD – 2003
6. R.A. MATA-TOLEDO, P. K. CUSHMAN Programmation SQL, (série Schaum's), la syntaxe normative des jointures, Ediscience 2003
7. Christian Soutou , De UML à SQL, Conception de bases de données, Eyrolles - Paris - juin 2002
8. Ryan K. Stephens, Ronald Plew, Conception de bases de données Collection " Référence" ouvrage a cheval entre l'analyse et la conception de bases de donnéesCampus Press – 2001
9. Serge Tahé : "introduction à la programmation Web avec Java" tutorial. Présente les bases de la programmation Web avec Java côté serveur avec les technologies des servlets, des pages JSP et XML. 215p. Sept 2002 http://tahe.developpez.com/java/web/
10. Emmanuel Puybaret :"Cahiers du Programmeur Java 1" 2003), et ensuite vous référer au tuto ici pour vous perfectionner : http://java.developpez.com/cours/#cours
11. David Menninger, Ventana Research." Pentaho4 Unites Enterprise Business intelligence and Data Integration.", Edition, June 22, 2011
12. Surya Mukherjee, Ovum. "Pentaho étend la couverture pour Big Data".Edition 8 mars 2012

13. A. Doucetand S.Gangarski. Entrepôts de données et Bases de Données multidimensionnelles, Chapitre12 du Edition: Bases de Données et Internet, Modèles, langages et systèmes. Editions Hermes, 2001.
14. Ravat, F. (2007). *Modèles et outils pour la conception et la manipulation de systèmes d'aide à la décision* (Doctoral dissertation, Université des Sciences Sociales-Toulouse I).
15. APROSI ., Rapport d'activité. http://www.APROSI..sn/sites/default/files/ RAPPORT%20%20D%27ACTIVITE%20FINAL.pdf Visité le 11/12/2018
16. Bouman, R., & Van Dongen, J. (2009). Pentaho solutions. *Business Intelligence and Data Warehousing with Pentaho and MYSQL*.
17. Golfarelli, M., & Rizzi, S. (2009). *Data warehouse design: Modern principles and methodologies* (Vol. 5, pp. 12-19). New York: McGraw-Hill.
18. Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record*, *26*(1), 65-74.

ANNEXE 1: OBJECT TABLE

```
/****** Object:  Table [dbo].[Fact_Energie]    Script Date: 03/19/2019 15:10:57 ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[Fact_Energie](
  [id_machine] [varchar](50) NULL,
  [id_local] [varchar](50) NULL,
  [id_entreprise] [varchar](50) NULL,
  [id_entrepot] [varchar](50) NULL,
  [id_capteur] [varchar](50) NULL,
  [id_temps] [varchar](50) NULL,
  [Energie_Consommee] [bigint] NULL
) ON [PRIMARY]
GO
SET ANSI_PADDING OFF
GO

/****** Object:  Table [dbo].[Dim_Machine]    Script Date: 03/19/2019 15:10:57 ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[Dim_Machine](
  [id_machine] [varchar](50) NULL,
  [nom_machine] [varchar](100) NULL,
  [marque] [varchar](50) NULL,
  [Nserie] [varchar](100) NULL,
) ON [PRIMARY]
GO
SET ANSI_PADDING OFF
GO


/****** Object:  Table [dbo].[Dim_LocalTechnique]    Script Date: 03/19/2019 15:10:57 ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
```

```
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[Dim_LocalTechnique](
  [id_local] [varchar](50) NULL,
  [nom_local] [varchar](100) NULL,
  [reference_local] [varchar](50) NULL,
  [reseau] [varchar](100) NULL,
  [type_local] [varchar](50) NULL
) ON [PRIMARY]
GO
SET ANSI_PADDING OFF
GO


/****** Object:  Table [dbo].[Dim_Entreprise]    Script Date: 03/19/2019 15:10:58 ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[Dim_Entreprise](
  [id_entreprise] [varchar](50) NULL,
  [nom_entreprise] [varchar](100) NULL,
  [reference] [varchar](50) NULL,
  [adresse] [varchar](100) NULL,
  [CodePostal] [varchar](50) NULL,
  [ville] [varchar](50) NULL,
  [localisation] [varchar](50) NULL,
  [code_responsable] [varchar](50) NULL,
  [actif] [varchar](50) NULL
) ON [PRIMARY]
GO
SET ANSI_PADDING OFF
GO


/****** Object:  Table [dbo].[Dim_Entrepot]    Script Date: 03/19/2019 15:10:58 ******/
SET ANSI_NULLS ON
GO
SET QUOTED_IDENTIFIER ON
GO
SET ANSI_PADDING ON
GO
CREATE TABLE [dbo].[Dim_Entrepot](
  [id_entrepot] [varchar](50) NULL,
  [nom_entrepot] [varchar](100) NULL,
```

```
    [reference_entrepot] [varchar](50) NULL,
    [type_entrepot] [varchar](100) NULL,
    [surface] [varchar](50) NULL,
    [nbr_niveau] [int] NULL,
    [gerant] [varchar](50) NULL
  ) ON [PRIMARY]
  GO
  SET ANSI_PADDING OFF
  GO


  /****** Object:  Table [dbo].[Dim_Capteur]  Script Date: 03/19/2019
15:10:58 ******/
  SET ANSI_NULLS ON
  GO
  SET QUOTED_IDENTIFIER ON
  GO
  SET ANSI_PADDING ON
  GO
  CREATE TABLE [dbo].[Dim_Capteur](
    [id_capteur] [varchar](50) NULL,
    [nom_capteur] [varchar](100) NULL,
    [unite] [varchar](50) NULL
  ) ON [PRIMARY]
  GO
  SET ANSI_PADDING OFF
  GO
```

RÉSUMÉ

Les nouvelles technologies de l'information de nos jours sont utilisés par les décideurs et dirigeants des entreprises pour obtenir une connaissance approfondie de l'entreprise et de définir leurs stratégies d'affaires, dans le but d'acquérir un avantage concurrentiel, d'améliorer la performance de l'entreprise, de répondre plus rapidement aux changements, d'augmenter la rentabilité, et d'une façon générale la création de valeur ajoutée de l'entreprise. Avec l'apparition des Data Warehouse, tous les utilisateurs peuvent désormais accéder facilement à l'information stratégique, ceci permet à l'entreprise d'être plus réactive dans sa prise de décisions. Ce mémoire traite l'étude et la conception d'un Data Warehouse pour le suivi de la consommation d'énergie dans le Domaine Industriel de Diamniadio, afin de permettre à l'APROSI de faire des analyses sur les résultats et prendre des décisions. L'approche retenue pour la réalisation de notre Data Warehouse est divisée en quatre phases principales: dans une première phase, nous avons effectué l'identification des besoins, une étude des outils techniques, ensuite une étude sur des différents outils d'implémentations d'entrepôt de données (SQLServer, Oracle, Pentaho, Talend Open Studio) et à la fin la conception du Data Warehouse.

Mots-clés : Data Warehouse, Diamniadio, ETL

ABSTRACT

New information technologies nowadays are used by decision-makers and business leaders to gain in-depth knowledge of the business and define their business strategies, in order to gain a competitive edge, improve the company's performance, to respond more quickly to changes, to increase profitability, and in a general way the creation of added value of the company. With the advent of the Data Warehouse, all users can now easily access strategic information, allowing the company to be more responsive in making decisions. This thesis deals with the study and design of a Data Warehouse for the monitoring of energy consumption in the Diamniadio Industrial Estate, in order to allow APROSI to analyze the results and make decisions. The approach adopted for the realization of our Data Warehouse is divided into four main phases: in a first phase, we carried out the identification of the needs, a study of the technical tools, then a study on various tools of implementations of data warehouse (SQLServer, Oracle, Pentaho, Talend Open Studio) and in the end the design of the Data Warehouse.

Keywords: Data Warehouse, Diamniadio, ETL