

Mémoire de fin d'études

Bac+5 en Ingénierie Informatique

Intitule 

**Mise en Œuvre d'un système
décisionnel pour la fonction énergie**

Présenté par:
Ayman BOUGADER

Encadré par:
Youness ETTAZARINI (ISMAGI)
Jaafar DEHBI (IT6)

Composition du Jury :

-
-
-

Président
Examinateur
Examinateur

Déclaration sur l'honneur anti-plagiat

Je certifie que le présent mémoire est uniquement et totalement le fruit d'un travail personnel et que toutes les sources auxquelles j'ai pu me référer sont clairement indiquées dans le corps du texte et figurent dans la bibliographie. Je suis conscient que le fait de ne pas citer une source ou de ne pas la citer clairement, correctement et complètement est constitutif de plagiat. Je déclare avoir pris connaissance des risques de sanctions administratives et disciplinaires encourues en cas de plagiat

Nom et prénom : Ayman BOUGADER

Date : 02/03/2020

Signature :



Dédicace

À ma chère Mère

*Ce travail est le fruit des sacrifices que tu as
Consentis pour mon éducation et ma formation.*

À mon cher Père

Aucune dédicace ne saurait exprimer l'amour et l'estime que j'ai toujours eus pour toi.

À mes frères Alae et Saad.

Votre amour et votre soutien me sont indispensables.

*À mes oncles et à mes tantes qui m'ont soutenu tout au long de mon parcours
scolaire.*

À mes meilleurs amis

Ainsi que leurs agréables parents.

*À toutes les merveilleuses rencontres que l'ISMAGI m'a permises le plaisir de
faire ces trois dernières années.*

Ayman Bouigader

Remerciements

Avant toute chose, je tiens à remercier chaleureusement M. Hicham ELACHGAR, et M. Youness RABBAB les directeurs généraux d'IT6, pour avoir accepté de m'intégrer au sein de la famille IT6.

J'adresse également mes sincères remerciements à M. Jaafar DEHBI **M**on encadrant au sein d'IT6 et M. Youness ETTAZARINI pour avoir accepté d'encadrer mon projet de fin d'études. Pour leur implication envers le suivi du déroulement du projet, ainsi que leurs interventions propices émanant aux moments adéquats de confusion ou de doute, qui m'aidaient à chaque reprise à retrouver le meilleur chemin à suivre.

Je tiens ensuite à exprimer ma profonde gratitude envers la Responsable du Pôle IT6 Academy, Mme. Btissam JABBAH, **P**our son accueil convivial.

A tous les membres du jury pour avoir accepté d'évaluer ce travail, je leur exprime ma grande reconnaissance.

Je remercie également toute l'équipe pédagogique de l'ISMAGI pour avoir assuré une formation de haute qualité.

Que tous ceux qui ont contribué de près ou de loin à l'aboutissement de ce projet trouvent l'expression de mes remerciements les plus sincères.

Résumé

Ce mémoire constitue une synthèse de mon projet de fin d'études effectué au sein d'IT6. Ce projet a pour objectif de mettre en œuvre une solution décisionnelle. Pour une meilleure conduite du projet, nous avons suivi le cycle de vie dimensionnel. Après avoir mis le projet dans son contexte, nous avons en premier lieu étudié les sources de données. Les différents besoins des utilisateurs ont été ensuite relevés, ce qui a permis d'aboutir à un dossier des spécifications fonctionnelles et un autre des spécifications techniques pour la solution. Nous avons alors enchaîné par la conception des Datamart constituant notre Datawarehouse et les tableaux de bord qui constitueront notre système de reporting. La réalisation de l'alimentation du Datawarehouse et celle des cubes OLAP ont été faite grâce à l'outil MSBI, et enfin une application web en ASP.net ChartsJS et RSTUDIO

Le résultat final de notre travail est une plateforme décisionnelle axée autour de thème Energie. Elle offre aux utilisateurs la possibilité d'accéder aux informations synthétisées et utiles, en consultant les Tableaux de bord

Mots-clés :

IT6, SIE, MSBI, Energie,

Datawarehouse, datamart, cube OLAP, tableau de bord, RStudio, Chartjs, ASP.net

Abstract

This document is a summary of my graduation project done within IT6. The objective of this project is to implement a decision-making solution. For a better management of the project, we followed the dimensional life cycle. After putting the project into context, we first studied the data sources. The different user needs were then identified, which resulted in a file of functional specifications and another of the technical specifications for the solution. We then continued by designing the datamarts constituting our datawarehouse and the dashboards which will constitute our reporting system. The datawarehouse and OLAP cubes were supplied with the MSBI tool, and finally a web application in ASP.net Chartjs and RSTUDIO. The end result of our work is a decision-making platform focused around the theme of Energy. It offers users the possibility of accessing synthesized and useful information, by consulting the Dashboards.

Keywords :

IT6, SIE, MSBI, Energy,
Datawarehouse, datamart, OLAP cube, dashboard, RStudio, Chartjs, ASP.net

Liste des abréviations

Abréviation	Désignation
BI	Business Intelligence
BPO	Business Process Outsourcing
SIE	System information energie
ETL	Extraction Transformation Loading
HTML	HyperText Markup Language
ITO	Information Technology Outsourcing
OLAP	On Line Analytical Processing
ROLAP	Relational On Line Analytical Processing
SCD	Slowly Changing Dimension
SGBD	Système de Gestion des Bases de Données
SQL	Structred Query Language
MAD	dirham marocain
MBA	Marge Brute d'Autofinancement
IP	Indice de Profitabilité
VAN	Valeur actuelle nette
TRI	Taux de Rentabilité Interne

Liste des figures

Figures 1: Pôles d'IT6	18
Figures 2: Références d'IT6	20
Figures 3: Schéma du cycle de vie dimensionnel	21
Figures 1: Diagramme de gantt	23
Figures 4: Le décisionnel au sein du Système d'information	27
Figures 5: Évolution des bases de données décisionnelles.	32
Figures 6: Comparaison entre OLAP et OLTP	33
Figures 7: L'approche d'Inmon (approche Top-Down)	35
Figures 8: L'approche de Kimball (ou approche Bottom-Up)	37
Figures 9: Inmon vs. Kimball : les caractéristiques majeures	37
Figures 10: Avantages et inconvénients (approches INMON & KIMBALL)	37
Figures 11: Typologie du projet	51
Figures 12: Profile du projet	51
Figures 13: Swot du Projet	52
Figures 14: Modèle en étoile	55
Figures 15: Datamart	57
Figures 16: Matrice des flux	60
Figures 17: Tableau des fonctions automatisées	61
Figures 18: Dictionnaires de données	62
Figures 19: Diagramme de cas d'utilisation	62
Figures 20: Diagrammes d'état transition	63
Figures 21: Conversion des données Csv Vers SQL Server	65
Figures 22: Table Dimension Comapany	66
Figures 23: Mapping Table Dimension Company	67
Figures 24: Table de fait	68
Figures 25: Lookup Categorie	68
Figures 26: Mapping table de fait	69
Figures 27: Connexion à la base de données	70
Figures 28: Schéma de source de donnée	71
Figures 29: Creation de cube	72
Figures 30: Deployment	72
Figures 31: Cube	73
Figures 32: Chargement des bibliothèque sur RSTUDIO	74
Figures 33: Tableaux de bord 'Dashbord'	77
Figures 34: Tableau de bord 'ENEXIS'	78
Figures 35: Tableau de bord 'LIANDER'	79
Figures 36: Tableau de bord 'LIANDER'	80
Figures 37: Tableau de bord 'RSTUDIO'	82
Figures 38: Formule de calcul de TRI	88

Liste des tableaux

Tableau 1 :	Livrables _____	24
Tableau 2 :	Indicateurs Consommation _____	47
Tableau 3 :	Indicateur Connection _____	48
Tableau 4 :	Indicateur Compteurs et livraison _____	48
Tableau 5 :	Axes d'analyses _____	49
Tableau 6 :	Matrice d'expression des besoins _____	56
Tableau 7 :	Tableaux de bord "Accueil" _____	58
Tableau 8 :	Tableaux de Bord des 'Sociétés ' _____	59
Tableau 9 :	Tableaux de bord "2T" _____	59
Tableau 10 :	Chargement des données sur RSTUDIO _____	74
Tableau 11 :	Traçage et visualisation _____	75
Détermination du cash – flow _____		86

Table des matières

I.	Chapitre I: Contexte général du projet	16
1.	Contexte général du projet	17
1.1	Organisme d'accueil: Présentation d'IT6.....	17
1.1.1	Pôles d'excellence.....	17
1.1.2	Références.....	20
2.	Présentation du projet	20
2.1	Motivation et Objectifs du Projets	20
2.2	Conduite du projet	21
3.	Conclusion	24
II.	Chapitre II: Etat de l'art	25
1.	Introduction.....	26
2.	Les systèmes décisionnels	26
2.1	La place du décisionnel dans l'entreprise.....	27
3.	Qu'est ce qu'un Data Warehouse.....	27
3.1	Quels types de données sont stockées dans une Datawarehouse ?.....	28
3.2	Quels sont les différents types de Data Warehouses ?	29
3.3	Comment et pourquoi les entreprises utilisent les Data Warehouses ?	30
3.4	Data Warehouse vs Data Lake : quelles sont les différences ?	30
3.5	Historique des Data Warehouse	31
4.	Etude Comparatif	32
4.1	OLAP VS OLTP.....	32
4.2	Système Centralisé	33
4.3	Système distribué.....	34
4.4	Différence entre système décisionnel et system distribué	34
5.	Modélisation et conception du Data Warehouse.....	35
5.1	L'approche d'Inmon (ou approche Top-Down).....	35
5.2	L'approche de Kimball (ou approche Bottom-Up)	36
5.3	Inmon vs. Kimball : les caractéristiques majeures	37
5.4	Avantages inconvenients des 2 approches.....	37

5.5	Comment bien choisir ?.....	38
5.6	Alimentation du Data Warehouse.....	38
5.6.1	Les phases de l'alimentation « E.T.L. »	38
5.6.2	L'extraction des données	39
5.6.3	La transformation des données.....	40
5.6.4	Le chargement des données	40
5.7	Politiques de l'alimentation.....	40
5.8	Les outils E.T.L.	41
5.9	Mise en œuvre du Data Warehouse	41
5.9.1	Requêtage ad-hoc :	41
5.9.2	Reporting :.....	42
5.9.3	Analyse dimensionnelle des données:.....	42
5.9.4	Tableaux de bord :	42
5.9.5	Data Mining :	43
6.	Maintenance et expansion.....	43
7.	Conclusion	44
III.	Chapitre III: Analyse et spécifications	45
1.	Étude de l'existant	46
1.1	Description de l'existant	46
2.	Recueil des besoins	46
2.1	Besoins principaux.....	46
3.	Cadrage fonctionnel du projet.....	47
3.1	Description des indicateurs	47
3.1.1	Indicateurs "énergie"	47
3.2	Description des axes d'analyses.....	49
4.	Cadrage technique	49
4.1	Macro Architecture	49
4.2	Architecture Applicative	50
4.3	Reflexion	50
4.4	Management du projet	51
4.4.1	Typologie.....	51
4.4.2	Profile.....	51

4.4.3	Swot	52
5.	Conclusion	52
IV.	Chapitre III: Conception.....	53
1.	Généralités sur le Datawarehouse.....	54
2.	Le modèle en étoile:.....	55
3.	Modélisation du Datawarehouse	55
3.1	Matrice d'expression des besoins	55
3.2	Schéma dimensionnel.....	56
3.3	Modélisation des datamarts	57
4.	Conception des tableaux de bord	57
3.1	Tableaux de bord 'Accueil'	58
3.2	Tableaux de bord "Société"	58
3.3	Tableaux de bord "2T"	59
3.4	Conception d'application WEB.....	59
3.4.1	Introduction	60
3.4.2	Choix de la méthodologie de conception	60
5.	Conclusion	63
V.	Chapitre IV: Mise en oeuvre.....	64
1.	Réalisation de l'ETL	65
1.1	Alimentation des tables de dimension	65
1.2	Alimentation de table de faits.....	67
2.	Cubes OLAP.....	70
2.1	Création des cubes OLAP	70
3.	Rstudio	73
3.1	Ggplot2	73
3.2	KableExtra	73
4.	Tableaux de Bord	75
3.1	Création des rapports	75
3.2	Outils de Travail.....	75
3.3	Restitution de donnée	77
3.3.1	Tableaux de Bord 'Dashbord'	77
3.3.2	Tableaux de bord 'ENEXIS'	78

3.3.3	Tableaux de bord 'LIANDER'	79
3.3.4	Tableaux de bord 'LIANDER'	80
3.3.5	Tableaux de bord 'RStudio'	81
5.	Conclusion	83
VI.	Analyse financière.....	84
1.	Evaluation financière	85
1.1	Valeur actuelle nette ou VAN.....	85
1.2	Indice de rentabilité IR ou indice de profitabilité IP	86
1.3	Taux de rentabilité interne ou TRI.....	88

Introduction générale

Pendant l'épidémie de Covid-19, le ministère de la Santé et IT6 ont décidé de suspendre indéfiniment le projet d'un système décisionnel, Mon encadrant a demandé un Prototype similaire.

Dans ce cadre, la base de données Dutch Energie a été présentée par le chef de projet cette Base de donnée son contexte est comme suit.

Enexis, Liander et Stedin sont les trois premières sociétés des Pays-Bas et, ensemble, elles fournissent de l'électricité et du Gaz à presque tout le pays. Chaque année, ils publient sur leur site internet un tableau des consommations d'énergie dans les zones sous leur administration.

Les données sont anonymisées en agrégant les codes postaux de sorte que chaque entrée décrit au moins 10 connexions.

Ce marché n'est pas concurrentiel, ce qui signifie que les zones sont attribuées. Cela signifie que chaque année, ils fournissent à peu près la même énergie aux mêmes codes postaux. De petits changements peuvent survenir d'une année à l'autre soit pour un changement de gestion, soit pour une agrégation différente des codes postaux.

C'est dans ce cadre nous avons lancé notre projet de fin d'étude pour s'offrir un outil de pilotage et d'aide à la décision. Notre projet consiste donc en la mise en œuvre d'une Plateforme décisionnel. Notre champ d'implication couvre le thème : Energie.

Ce mémoire présente les différentes étapes de réalisation de notre projet. Il se compose de quatre chapitres dont la description est comme suit.

Le premier chapitre aborde le contexte général du projet. Puis expose la motivation et les objectifs du projet. Ensuite, on passe à la démarche suivie, le planning avant de terminer livrables attendus .

Le deuxième chapitre est dédié à l'analyse et spécifications des besoins. Ce qui se traduit par une étude de l'existant et ses limites et le recensement des différents besoins de la direction pour arriver à un cadrage fonctionnel et technique de la solution proposée.

Quant au troisième chapitre, il traite la conception du modèle multidimensionnel de notre Datawarehouse et d'application Web Contenant notre Tableaux de Bord.

Le quatrième chapitre détaille la phase de la mise en œuvre du projet dans ses différentes étapes à savoir l'alimentation du Datawarehouse et la création des cubes OLAP, Application Web et Tableaux de bord. Pour clôturer ce mémoire, nous résumons les résultats et acquis réalisés lors de ce projet, les difficultés rencontrées et des perspectives à ce travail.

I. Chapitre I: Contexte général du projet

Ce chapitre a comme objectif de décrire l'organisme d'accueil, et le contexte général du projet, qui consiste en la mise en œuvre d'un System décisionnel pour L'étude et l'analyse de la consommation d'énergie au Pays-Bas. Ainsi que nous présenterons ensuite notre projet et ses objectifs. Enfin, nous décrirons la démarche suivie, le planning et le plan assurance qualité adoptés, et les livrables attendus.

1. Contexte général du projet

1.1 Organisme d'accueil: Présentation d'IT6

En plus d'être un opérateur de formation confirmé, IT6 est aussi un acteur majeur sur le marché du conseil en gouvernance, management et systèmes d'information, qui se distingue par une approche différente de la performance et une combinaison d'expertise à forte valeur ajoutée.

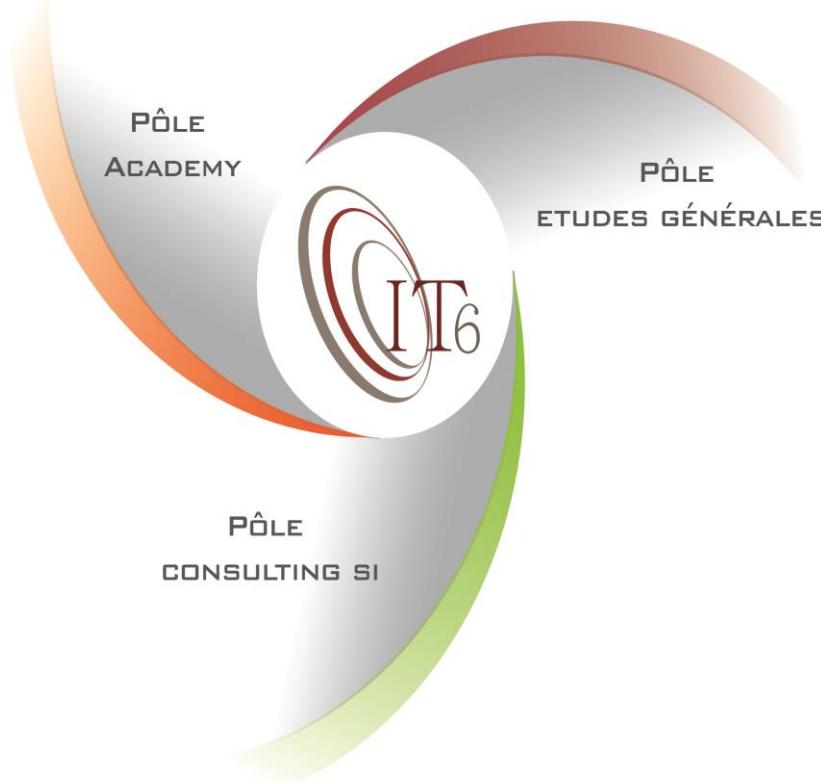
- IT6 a comme mission d'accompagner, en partant de la vision stratégique à la mise en œuvre opérationnelle des processus de management et des systèmes d'information.
- Leur Objectif c'est le développement et création de la valeur, performance au service de tous.
- Leur Savoir-faire c'est d'Optimiser vos projets et proposer des solutions durables, à même de répondre aux besoins évolutifs de votre stratégie métier.

IT6 a une gamme complète de prestations d'audit, conseil et formations est conçue pour apporter une réponse adaptée aux besoins, et constitue le fer de lance de la réussite des organisations.

1.1.1 Pôles d'excellence

Forts ~~de~~ d'expérience sur de nombreuses missions, l'approche d'IT6 tient compte des spécificités propres à chaque organisation.

Ils portent la plus extrême attention au changement jusqu'à l'obtention des résultats concrets.



Figures 1: Pôles d'IT6

Pour servir les projets, IT6 constitue une équipe de consultants riches d'une double compétence technique et fonctionnelle.

Investis pleinement au cœur des équipes de clients , ils comprendront leur enjeux pour accélérer leur réussite, en s'appuyant sur les valeurs qui constituent leur engagement.

Pour la mise en place des conditions nécessaires à une croissance durable pour les entreprises, IT6 s'engage et accompagne les clients dans l'établissement de stratégies globales pour réussir leurs projets.

1.1.1.1 *Pôle Academy*

Dans ce Pôle d'excellence, IT6 offre des formations animées par les meilleurs spécialistes marocains et internationaux et organise des séminaires qui dressent l'État de l'art, analysent les nouvelles tendances et constituent des références dans les principaux domaines de l'informatique.

Destinés aux DSI ET à leurs adjoints, aux chefs de projets, aux ingénieurs d'études et aux consultants, les séminaires de IT6 se sont forgés une réputation de sérieux, de rigueur et de qualité aujourd'hui.

1.1.1.2 Pôle consulting SI

Dans le respect des exigences métier et des culture d'entreprise, IT6 met en œuvre son savoir-faire pour proposer une Gouvernance de SI, à même de répondre, de façon cohérente, à l'ensemble des besoins des clients :

- ⇒ Accompagnement à la mise en place d'un système de management de la sécurité de l'information.
- ⇒ Assistance à la mise en place du Help Desk, SLAs...
- ⇒ Audit de la sécurité des systèmes d'information, Audit technique, test d'intrusion interne et externe...

Parmi les véritables accélérateurs d'IT6 de performances du SI, Ils propose les offres suivantes:

- ⇒ Élaboration des schémas directeurs informatiques ;
- ⇒ Audit et évaluation des SI ;
- ⇒ Audit et mise à niveau de la sécurité SI ;

Assistance à maîtrise d'ouvrage dans les projets en SI

1.1.1.3 Pôle études générales

Les consultants d'IT6 sont unis autour de valeurs assurant la poursuite des mêmes objectifs et la cohérence des interventions chez leur clients.

Grâce à une longue expérience et une maîtrise de toutes les techniques analytiques et décisionnelles, ils interviennent en proposant une approche globale des problématiques, et en apportant un savoir-faire managérial, organisationnel et technique en matière :

- de conseil stratégique et organisationnel,
- d'études sectorielles,
- d'études économiques ou sociaux,
- de mise en place de dispositifs de gouvernance et de pilotage des performances...

1.1.1.4 Vision

IT6 est l'un des rares cabinets au Maroc à être véritablement spécialisé dans le conseil en management des systèmes d'information et à avoir développé une offre spécifique pour les besoins des entreprises.

IT6 vise sa contribution en acteur de premier plan au niveau national en alimentant en permanence sa force d'innovation et de qualité.

1.1.2 Références

Les clients d'IT6 font partie aussi bien du secteur privé que du secteur public et nous prestations couvrent les organisations de la multinationale à la PME.



Figures 2: Références d'IT6

2. Présentation du projet

2.1 Motivation et Objectifs du Projets

Le but de notre Projet est de définir les indicateurs et de générer les tableaux de bord et les rapports qui offrent une meilleure visibilité et qui mettent à disposition l'information essentielle pour la prise de décision relativement au thème énergie.

Les tableaux de bord à générer visent :

Le suivi des énergies afin de mesurer l'évolution de la Consommation dans les deux catégories électricité et gaz.

Le suivi de l'évolution de l'utilisation des compteurs intelligents.

Le suivi d'augmentation du nombre de connections.

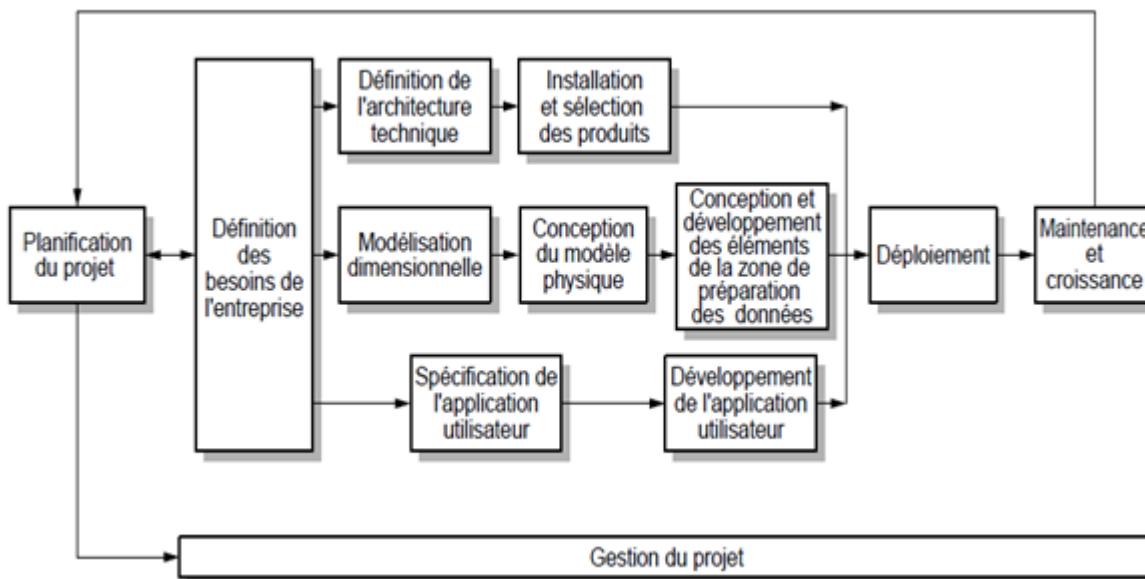
Le suivi de l'évolution de consommation pendant les heures à tarif réduit. ;

2.2 Conduite du projet

2.2.1 Démarche

Notre projet étant à caractère décisionnel, pour sa conduite, nous allons suivre le cycle de vie dimensionnel. Ce cycle commence par la planification du projet puis passe à la définition des besoins qui constituent le point de départ de trois trajectoires parallèles qui sont la technologie, les données et l'interface utilisateur. Ces trajectoires se rejoignent en phase de déploiement qui est suivie par la phase de maintenance et croissance.

Le schéma ci-dessous représente la succession des tâches de haut niveau nécessaires à la conception,



Figures 3: Schéma du cycle de vie dimensionnel

Les étapes de la figure ci-dessus sont :

Planification du projet : aborde la définition et l'étendue du projet. Elle permet l'affectation des tâches à leurs durées et à leur séquencement.

Définition des besoins : permet d'identifier les besoins de l'entreprise et détermine les données requises pour répondre aux besoins d'analyse des utilisateurs.

Modélisation dimensionnelle : commence par la construction d'une matrice qui représente les processus positions clé et leurs dimensionnalités. A partir de cette matrice, on élabore une analyse plus détaillée des données des systèmes sources, puis on développe le modèle dimensionnel.

Conception du modèle physique : définit les structures nécessaires pour l'implémentation du modèle dimensionnel. Cette étape nécessite la détermination des règles de nommage des objets et la mise en place de l'environnement de la base de données.

Conception et le développement de la zone de préparation des données : se déroule en trois phases majeures : Extraction, Transformation et Chargement.

Définition de l'architecture technique : permet d'avoir une vision globale de l'architecture technique à mettre en œuvre. Dans cette phase, trois facteurs doivent être pris en compte : les besoins, l'environnement existant et les orientations techniques stratégiques planifiées.

Sélection des produits et l'installation  permet la sélection des composants spécifiques tels que la plateforme matérielle, le SGBD (Système de Gestion de Bases de Données), et les outils de préparation et d'accès aux données. Une fois évalués et sélectionnés, ceux-ci devront être installés et testés minutieusement.

Spécifications de l'application utilisateur : décrit les maquettes d'états, les critères de sélection laissés à l'utilisateur et les calculs nécessaires.

Développement de l'application utilisateur : commence par une compréhension commune par l'équipe de développement et les utilisateurs finaux des spécifications établies.

Déploiement : est le point de convergence de la technologie, des données et des applications utilisateur. Une formation des utilisateurs est nécessaire. Elle intègre tous les aspects de cette convergence avant de permettre à ceux-ci d'accéder à l'application. La prise en compte des demandes d'évolution et de correction est également indispensable.

Maintenance et croissance : permet de s'occuper des utilisateurs en leur procurant un service de support et une formation continue.

Gestion du projet : garantit que les activités du cycle de vie dimensionnel restent sur la bonne voie et sont bien synchronisées.

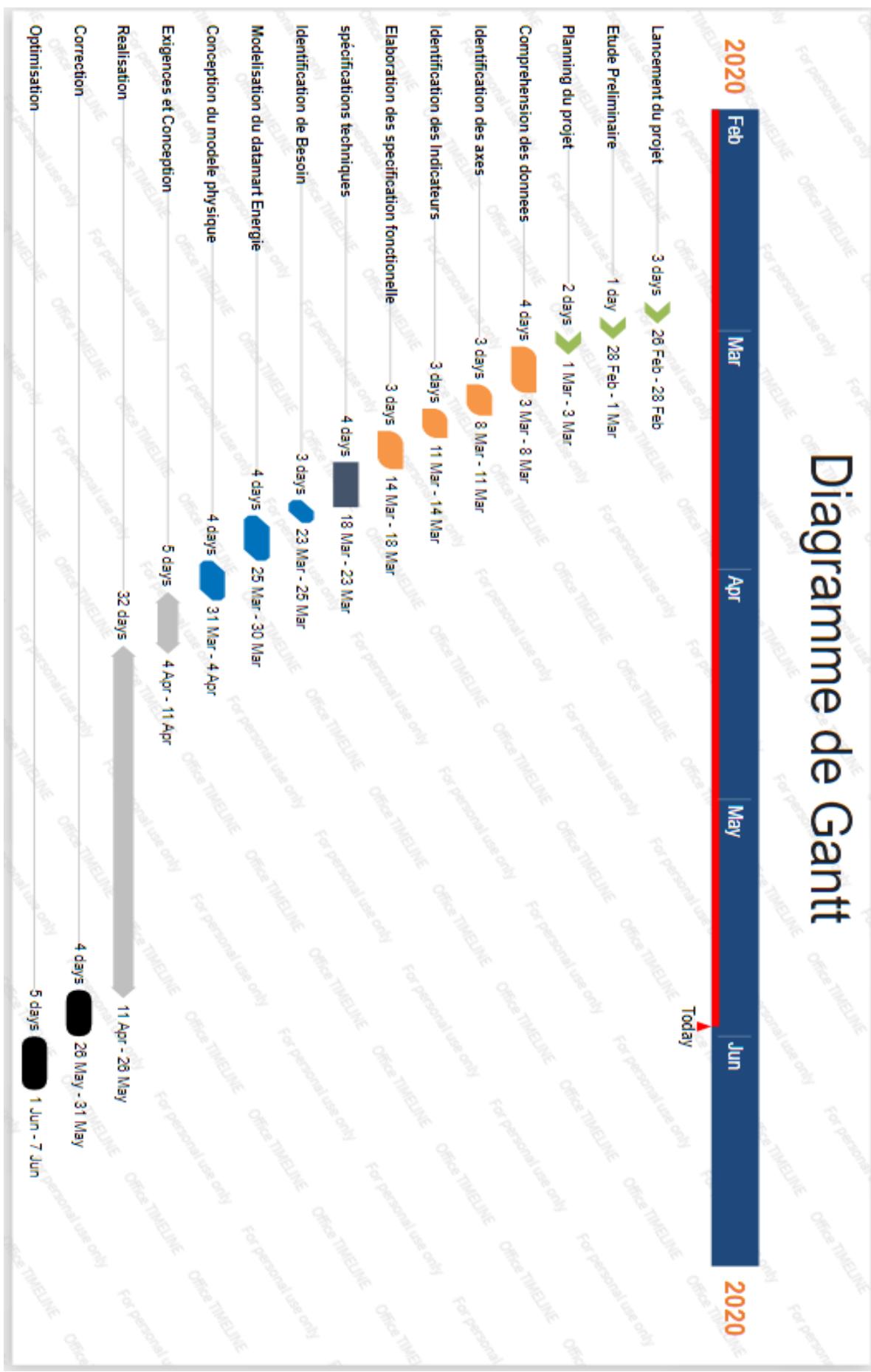
2.2.2 *Planning*

La planification du projet découle du cycle de vie dimensionnel, décrit auparavant, afin de placer les éléments du projet dans le bon ordre au bon moment

Le projet est découpé en phases chronologiques et organisationnelles, qui elles aussi sont découpées en tâches pour faciliter l'estimation du temps de réalisation du projet

La répartition des tâches à accomplir, leurs durées et leur succession sont décrites dans le diagramme de GANTT présenté ci-dessous.

Diagramme de Gantt



Figures 1: Diagramme de gantt

2.2.3 *Livrables*

Pour chaque étape du diagramme de GANTT, une remise de livrables est nécessaire, afin d'assurer la traçabilité du projet et de bien mener la communication avec les utilisateurs pour pouvoir valider chaque étape. Le tableau ci-dessous montre les livrables de chaque phase.

Phase	Livrables
Planification du projet	Lancement, Etude Préliminaire, Planning.
Etude et définitions des besoins	Compréhension des données, Identification du besoin Identification des axes, Identification des indicateurs
Architecture technique	Spécification fonctionnelle Spécification technique.
Modélisation multidimensionnelle	Modélisation du Datamart
Conception de l'application utilisateur	Conception du modèles physique , Exigences et conception, Réalisation
Maintenance et croissance	Correction, Optimisation.

Tableau 1 : *Livrables*

3. Conclusion

Ce premier chapitre met en évidence le contexte général du projet sur lequel s'articule notre Projet de Fin d'Etudes chez IT6 et qui traite la mise en œuvre d'une plateforme décisionnel concernant l'énergie.

Ce projet consiste donc à la conception et l'alimentation d'un Datamart pour le SIE et L'application web contenant les tableaux de bord

Tout d'abord, l'organisme d'accueil et son client ont été présentés afin de situer le cadre général du projet.

Enfin, la conduite du projet a été établie en choisissant la démarche cycle de vie dimensionnel et en élaborant le planning des différents lots liés au projet.

Il convient maintenant d'attaquer le premier volet cité dans notre démarche à savoir l'analyse et spécifications des besoins ce qui est l'objet du chapitre suivant.

II. Chapitre II: Etat de l'art



1. Introduction

Toutes les entreprises du monde disposent d'une masse de données plus ou moins considérable. Ces informations proviennent soit de sources internes (générées par leurs systèmes opérationnels au fil des activités journalières), ou bien de sources externes (web, partenaire, .. etc.).

Cette surabondance de données, et l'impossibilité des systèmes opérationnels de les exploiter à des fins d'analyse conduit, inévitablement, l'entreprise à se tourner vers une nouvelle informatique dite décisionnelle qui met l'accent sur la compréhension de l'environnement de l'entreprise et l'exploitation de ces données à bon escient. En effet, les décideurs de l'entreprise ont besoin d'avoir une meilleure vision de leur environnement et de son évolution, ainsi, que des informations auxquelles ils peuvent se fier. Cela ne peut se faire qu'en mettant en place des indicateurs « business » clairs et pertinents permettant la sauvegarde, l'utilisation de la mémoire de l'entreprise et offrant à ses décideurs la possibilité de se reporter à ces indicateurs pour une bonne prise de décision.

Le « Data Warehouse », « Entrepôt de données » en français, constitue, dans ces conditions, une structure informatique et une fondation des plus incontournables pour la mise en place d'applications décisionnelles.

Le concept de Data Warehouse, tel que connu aujourd'hui, est apparu pour la première fois en 1980 l'idée consistait alors à réaliser une base de données destinée exclusivement au processus décisionnel. Les nouveaux besoins de l'entreprise, les quantités importantes de données produites par les systèmes opérationnels et l'apparition des technologies aptes à sa mise en œuvre ont contribué à l'apparition du concept « Data Warehouse » comme support aux systèmes décisionnels.

2. Les systèmes décisionnels

La raison d'être d'un entrepôt de données, comme évoqué précédemment, est la mise en place d'une informatique décisionnelle au sein de l'entreprise. Pour cela il serait assez intéressant de définir quelques concepts clés autour du décisionnel.

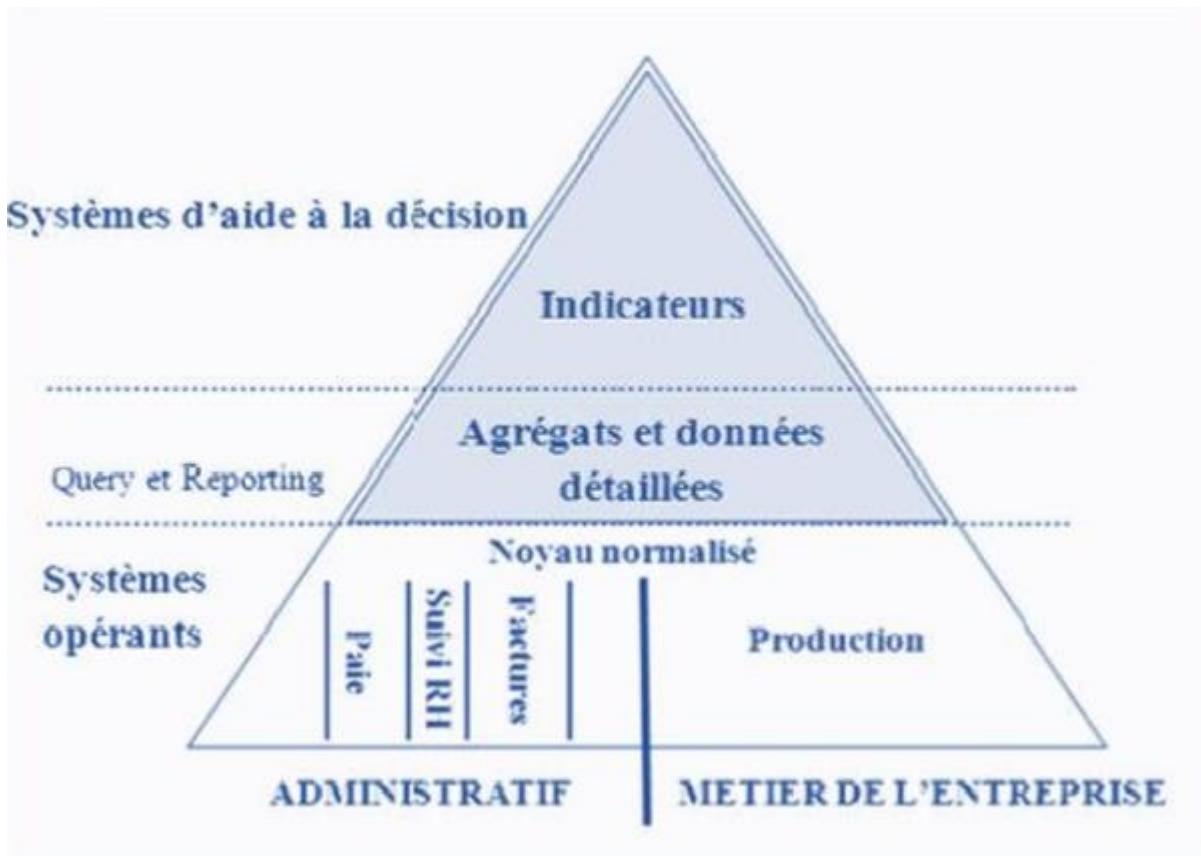
Afin de mieux comprendre la finalité des systèmes décisionnels, nous nous devons de les placer dans leurs contextes et rappeler ce qu'est un système d'information.

«Le système d'information est l'ensemble des méthodes et moyens de recueil de contrôle et de distribution des informations nécessaires à l'exercice de l'activité en tout point de l'organisation. Il a pour fonction de produire et de mémoriser les informations, de l'activité du système opérant (système opérationnel), puis de les mettre à disposition du système de décision (système de pilotage)». Les différences qui existent entre le système de pilotage et le système opérationnel, du point de vue fonctionnel ou des tâches à effectuer, conduit à l'apparition des « systèmes d'information décisionnels ». Ces différences seront clairement illustrées un peu plus loin dans notre document.

Les différences qui existent entre le système de pilotage et le système opérationnel, du point de vue fonctionnel ou des tâches à effectuer, conduit à l'apparition des « systèmes d'information décisionnels » (S.I.D.). Ces différences seront clairement illustrées un peu plus loin dans notre document. Les origines des SID remontent au début de l'informatique et des systèmes d'information qui ont, tous deux, connu une grande et complexe évolution liée notamment à cette évolution. Cette évolution se poursuit à ce jour. Parmi les différentes définitions du décisionnel données on trouve : "Le Décisionnel est le processus

visant à transformer les données en informations et, par l'intermédiaire d'interrogations successives, transformer ces informations en connaissances."

2.1 La place du décisionnel dans l'entreprise



Figures 4: Le décisionnel au sein du Système d'information

3. Qu'est ce qu'un Data Warehouse

Les Datawarehouse ou entrepôts de données sont des bases de données permettant de stocker des données historiques structurées non volatiles orientées sujet afin de les analyser. Découvrez la définition, les avantages, et les cas d'usage des Datawarehouse.

Une Datawarehouse est une base de données relationnelle pensée et conçue pour les requêtes et les analyses de données, la prise de décision et les activités de type Business Intelligence davantage que pour le traitement de transactions ou autres usages traditionnels des bases de données.

Les informations stockées dans la Datawarehouse sont historiques, et offrent une vue d'ensemble des différentes transactions qui ont eu lieu au fil du temps. Les données redondantes sont souvent incluses dans les Datawarehouse pour offrir aux utilisateurs plusieurs vues sur l'information. C'est la raison pour laquelle les données stockées dans la Datawarehouse sont souvent agrégées pour permettre aux utilisateurs d'y accéder plus facilement.

En plus d'une base de données relationnelle, un environnement Datawarehouse intègre un outil d'extraction, de transport, de transformation et de chargement de données (ETL). On retrouve aussi un moteur de traitement analytique en ligne (OLAP), des outils d'analyse client, et d'autres applications permettant de gérer le traitement des données collectées.

L'une des principales particularités d'une Datawarehouse est que les informations y sont classées par sujets (clients, produits...). De fait, ce qui définit réellement une Datawarehouse est le type de données qu'elle contient et les personnes qui l'utilisent.

3.1 Quels types de données sont stockées dans une Datawarehouse ?

Les données stockées dans une Data Warehouse présentent plusieurs spécificités. Elles sont séparées des systèmes opérationnels, mais aussi accessibles et disponibles pour les requêtes.

Ces données sont orientées sujet, et intégrées. Elles sont aussi nommées et définies de façon consistante. Par ailleurs, ces données sont associées à des périodes de temps définies.

Enfin, elles sont statiques (non volatiles), ce qui signifie qu'aucune mise à jour n'est effectuée sur ces données.

Les quatre caractéristiques des Data Warehouses selon leur inventeur William H. Inmon Selon William H. Inmon, l'inventeur du terme, les Data Warehouses présentent quatre caractéristiques spécifiques. Elles doivent être orientées sujet, intégrées, non volatiles et « time-variant ».

Les Data Warehouses doivent être orientées sujet, ce qui signifie qu'il doit être possible de les définir par leur sujet. Par exemple, une warehouse peut être déployée spécialement pour analyser les données liées aux ventes de l'entreprise. Cette Data Warehouse servira à répondre à des questions comme « quels ont été les meilleurs clients pour tel produit au cours de l'année précédente ».

Dans la même logique, la Data Warehouse doit être en mesure d'assembler des données en provenance de différentes sources dans un format consistant. Elles doivent permettre de résoudre les problèmes comme les conflits de noms et les incohérences en termes d'unités de mesure. On parle là d'intégration.

Troisièmement, les Data Warehouses doivent être non-volatiles. Cela signifie qu'une fois qu'une donnée est entrée dans la Warehouse, elle ne doit plus changer. L'utilisateur est ainsi en mesure d'analyser les données telles qu'elles ont été stockées dans la Warehouse.

Dernière caractéristique des Data Warehouses, elles doivent être time-variant. Celà signifie qu'elles permettent de focaliser les analyses sur les changements survenus au fil du temps à partir de larges ensembles de données, afin de découvrir des tendances. C'est ce qui oppose les Data Warehouses aux systèmes OLTP dont les données opérationnelles sont atomiques et ne reflètent que la valeur actuelle de la dernière transaction.

3.2 Quels sont les différents types de Data Warehouses ?

On distingue quatre types de Data Warehouses : les systèmes de gestion de base de données (SGBD) relationnelles traditionnels, les SGBD relationnelles spécialisés, les Data Warehouses Appliances et les Data Warehouses Cloud. Les SGBD sont les Data Warehouses les plus courantes.

Dans la plupart des cas, on utilise un SGBD relationnel, mais il convient de noter que n'importe quel type de SGBD peut être utilisé. Dans la plupart des cas, un SGBD de Data Warehouse propose plusieurs fonctionnalités et caractéristiques additionnelles permettant de les utiliser efficacement pour les activités et opérations de Data Warehousing. Une plateforme Data Warehouse peut aussi embarquer des logiciels additionnels. On parle là de SGBD de vendeurs comme IBM, Microsoft, Oracle ou SAP.

Les SGBD spécialisées quant à elles se distinguent des SGBD traditionnels par la façon dont elles ont été améliorées pour pouvoir prendre en charge les workloads Data Warehouse. Ces produits sont proposés par des vendeurs comme HP et SAP.

Les Datawarehouse Appliance sont conçues pour délivrer des services de Data Warehousing clé en main directement. En général, ces produits combinent logiciel et matériel avec une SGBD relationnelle ou analytique préinstallée et configurée sur le hardware requis.

Le serveur est configuré avec la bonne quantité de mémoire et de stockage, et le SGBD est installé par le vendeur. L'utilisateur se contente de brancher le matériel et de l'allumer. On peut citer comme exemple les produits de Teradata, IBM et Oracle.

Dans la dernière catégorie, celle des Data Warehouses Cloud, le SGBD n'est pas installé sur site. L'utilisateur y accède via internet. On parle là de « Data Warehouse en tant que Service », de la famille des XaaS. L'utilisateur n'a pas besoin de disposer d'un SGBD ou de matériel dédié sur site. Les leaders du marché des Data Warehouses cloud sont Microsoft, Amazon et IBM.

Précisons que l'on compte aussi de plus en plus de solutions hybrides intégrant diverses capacités de stockage et d'accès aux données pour les données structurées et non structurées. On peut citer les vendeurs Actian et Pivotal.

Par ailleurs, ces différents types de plateformes Data Warehouses peuvent être déployés de plusieurs façons : Enterprise Data Warehouse (EDW), Data Mart, ou une combinaison des deux.

Une EDW est utilisée à l'échelle d'une entreprise entière, tandis qu'un Data Mart est plus petit et focalisé sur les besoins individuels ou spécifiques à un département. De même, en fonction de la taille de l'entreprise, il est possible de créer plusieurs Datamarts et les intégrer avec une EDW.

3.3 Comment et pourquoi les entreprises utilisent les Data Warehouses ?

Grâce à l'Online Analytical Processing (OLAP), les entreprises sont en mesure de dégager des insights de leurs opérations par le biais d'un accès interactif et itératif aux données stockées. Ceci permet aux responsables des entreprises d'améliorer les prises de décisions en effectuant des requêtes pour examiner les processus, les performances et les tendances de leurs entreprises.

Une Data Warehouse peut être utilisée pour suivre, gérer et améliorer les performances d'une entreprise. Elle peut être utilisée pour suivre et modifier une campagne marketing. On peut s'en servir pour passer en revue et optimiser la logistique et les opérations, ou pour améliorer l'efficacité du développement de produit.

Les entreprises utilisent aussi les Data Warehouses pour lier et accéder aux informations en provenance de sources multiples. Ces solutions permettent aussi de gérer et d'améliorer les relations clients. Les Data Warehouses peuvent permettre de prédire les futures tendances et besoins, et enfin d'améliorer la qualité des données.

III.5 Data Warehouses : quels sont les avantages ?

Les Data Warehouses présentent de nombreux avantages. Pour les responsables informatiques, elles permettent notamment de séparer les processus analytiques des processus d'exploitation pour améliorer les performances dans ces deux domaines.

Pour les entreprises, une plateforme Data Warehouse est une façon pratique de visualiser le passé sans affecter les opérations quotidiennes. En effectuant des requêtes et des analyses de données au sein de la Data Warehouse, les entreprises peuvent améliorer leurs opérations et leur efficience, et ainsi augmenter leurs revenus et leurs bénéfices.

3.4 Data Warehouse vs Data Lake : quelles sont les différences ?

Les Data Warehouses sont utilisées depuis près de 30 ans. Depuis peu toutefois, les Data Lakes gagnent en popularité à tel point que certains pensent qu'ils vont remplacer les Warehouses. En réalité, il convient de garder en tête que ces deux types de systèmes présentent d'importantes différences et ne sont pas utilisés de la même façon.

Une Data Warehouse rassemble une grande quantité de données accumulées au sein d'une entreprise en provenance de différentes sources de données. Elle est utilisée pour prendre des décisions. De son côté, le Data Lake est une banque de stockage servant à contenir une immense quantité de données brutes dans leur format d'origine jusqu'à ce que l'entreprise en ait besoin. Data Lakes et Data Warehouses se distinguent sur plusieurs points : les données, le traitement des données, le stockage, l'agilité, la sécurité et les utilisateurs.

Les Data Warehouses ne peuvent accueillir que des données structurées. Un Data Lake est capable de stocker des données structurées, semi-structurées, ou non structurées. En ce qui concerne le traitement des données, les données chargées dans une Data Warehouse doivent passer par l'étape du schema-on-write qui consiste à leur conférer une forme et une structure (un modèle). Dans le cas des Data Lakes, les données sont stockées sous leur forme brute et l'utilisateur leur donne forme en cas de besoin. C'est ce qu'on appelle le schema-on-read.

Le stockage dans une Data Warehouse peut être cher, surtout si le volume de données est large. Le stockage sur Data Lake revient souvent moins cher, car la plupart des technologies Big Data reposent sur des logiciels open source conçus pour être installés sur du matériel lowcost.

Une Data Warehouse est une banque de données structurée. Il n'est donc pas difficile techniquement de changer la structure. Cependant, ce processus peut prendre du temps en fonction des business processes qui y sont attachées. Contrairement à la Data Warehouse, le Data Lake n'a pas de structure. De fait, les Data Developers et Data Scientists peuvent aisément configurer et reconfigurer les modèles de données, les requêtes et les applications. Les Data Warehouses sont donc moins agiles.

En revanche, en termes de sécurité, les Data Warehouses profitent de leur ancienneté et de leur maturité. Les données sont moins sécurisées au sein d'un Data Lake, même si des progrès sont effectués dans ce domaine. La dernière différence entre Data Warehouses et Data Lakes sont les utilisateurs auxquels ils se destinent. Les Data Warehouses sont principalement utilisées par les responsables d'entreprises, tandis que les Data Lakes sont généralement utilisés par des Data Scientists dans des secteurs scientifiques.

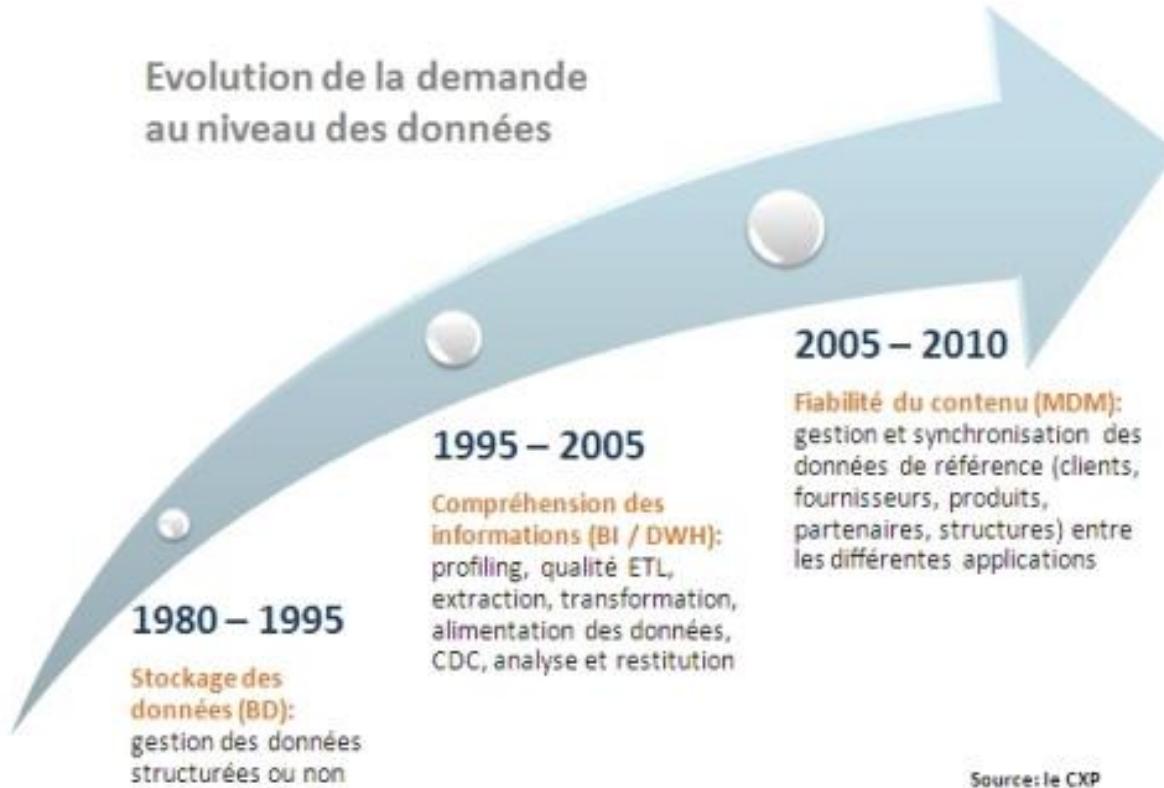
3.5 Historique des Data Warehouse

L'origine du concept « Data Warehouse » D.W (entrepôt de données en français) remonte aux années 80, durant lesquelles un intérêt croissant au système décisionnel a vu le jour, dû essentiellement à l'émergence des SGBD relationnel et la simplicité du modèle relationnel et la puissance offerte par le langage SQL,

Au début, le Data Warehouse n'était rien d'autre qu'une copie des données du système opérationnel prise de façon périodique, dédiée à un environnement de support à la prise de décision. Ainsi, les données étaient extraites du système opérationnel, stockées dans une nouvelle base de données «concept d'infocentre », le motif principal étant de répondre aux requêtes des décideurs sans pour autant altérer les performances des systèmes opérationnels.

Le Data Warehouse, tel qu'on le connaît actuellement, n'est plus vu comme une copie ou un cumul de copies prises de façon périodique- des données du système opérationnel. Il est devenu une nouvelle source d'information, alimenté avec des données recueillies et consolidées des différentes sources internes et externes.

Evolution de la demande au niveau des données



Figures 5: Évolution des bases de données décisionnelles.

4. Etude Comparatif

Nous voulons ici effectuer une étude comparative sur différents Méthode et système et bases de données . L'objectif de notre recherche empirique sera multiple. Tenterons de montrer l'importance de l'informatique décisionnel les concepts et les méthodes liée.

4.1 OLAP VS OLTP

OLTP : (*On-Line Transaction Processing*)

- Sont les outils traditionnels de gestion et d'exploitation des données sont du type transactionnel Le transactionnel est basé sur un mode d'exploitation de données axé la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données.
- Le système transactionnel est généralement une application avec une BD, stockant les données courantes d'une organisation (Pb de gestion automatique des archives dans les systèmes transactionnels)

OLAP : (*Online Analytical Processing*)

Le OLAP, ou Online Analytical Processing, est une technologie de traitement informatique (computer processing). Elle permet à un utilisateur de consulter et d'extraire facilement les données pour les comparer de différentes façons. C'est un outil inscrit dans analysis services d'aide à la décision bien pratique pour une entreprise. Les données OLAP sont stockées sur une base multidimensionnelle, aussi appelées Cubes OLAP, pour faciliter ce type d'analyses. Un serveur OLAP est nécessaire.

OLTP	OLAP
Orienté transaction	Orienté analyse
Orienté application	Orienté sujet
Données courantes	Données historisées
Données détaillées	Données agrégées
Données évolutives	Données statiques
Utilisateurs nombreux, administrateurs/opérationnels	Utilisateurs peu nombreux, manager
Temps d'exécution: court	Temps d'exécution: long

Figures 6: Comparaison entre OLAP et OLTP

4.2 Système Centralisé

L'informatique centralisée ou architecture centralisée regroupe les ressources nécessaires au traitement sur un hôte central, généralement à l'aide de terminaux connectés à un ordinateur central. Cet ordinateur peut contrôler directement tous les périphériques (s'ils y sont physiquement connectés) ou ils peuvent être gérés par un serveur de terminaux. Si les terminaux ont la capacité, ils peuvent être en mesure de se connecter à l'ordinateur central directement à partir d'un réseau informatique. L'ordinateur central peut être un serveur central dans le cas d'une architecture client-serveur. Les terminaux peuvent être de type passif ou actif, ils peuvent afficher des environnements CLI, TUI ou GUI, il peut s'agir d'une console système, d'une imprimante pour un client léger ou d'un plus gros client par exemple.

4.3 Système distribué

L'architecture distribuée ou l'informatique distribuée désigne un système d'information ou un réseau pour lequel l'ensemble des ressources disponibles ne se trouvent pas au même endroit ou sur la même machine. Ce concept, dont une version peut être une combinaison de transmissions du type client-serveur

4.4 Différence entre système décisionnel et system distribué

Autorité

systèmes centralisés se concentrent sur le niveau de la gestion , ce qui lui donne plus d'autorité pour prendre des décisions et de les relayer aux employés. Les systèmes centralisés peuvent donc devenir autoritaire , car ils découragent l'innovation par d'autres employés . Systèmes distribués ou décentralisés , en revanche , sont plus ouverts à la délégation. Ils se partagent le pouvoir et les capacités de prise de décision plus uniformément dans toute l'organisation .

Interaction

Interaction entre l'équipe de direction et d'autres parties de l'organisation est limitée dans un système centralisé . Les gestionnaires sont les seuls décideurs, de sorte qu'ils ont tendance à être trop occupé pour interagir avec le personnel autre que lorsque le besoin s'en fait sentir. Cependant, pour les systèmes décentralisés, toutes les parties de l'organisation sont libres d'interagir parce que les gestionnaires ont moins de responsabilités et le processus décisionnel implique d'autres membres du personnel.

Applicabilité

les systèmes centralisés sont les mieux adaptés aux petites entreprises, où le contrôle global est plus facile parce que le marché du travail est plus faible. Les entreprises franchisées, d'autre part, ne peuvent pas utiliser les systèmes centralisés. Bien que les décisions de développement de marque, le contrôle qualité et la commercialisation d'autres produits proviennent de la haute direction, les propriétaires de franchises sont généralement libres de gérer l'entreprise leur chemin tant qu'ils respectent les normes établies. Par conséquent, les grandes industries et les entreprises adoptent souvent une gestion décentralisée

Communication

communication pour un système centralisé est de haut en bas ; . Informations sur l'organisation découle de la plus - niveau supérieur vers le niveau le plus bas des salariés. Cependant, il est généralement possible de décisions par les gestionnaires de niveau moyen décision. Pour les systèmes décentralisés, la communication est plus arrondie . L'information et la prise de décision dans l'entreprise peuvent circuler de haut en bas, de bas en haut, ainsi que l'ensemble des ministères.

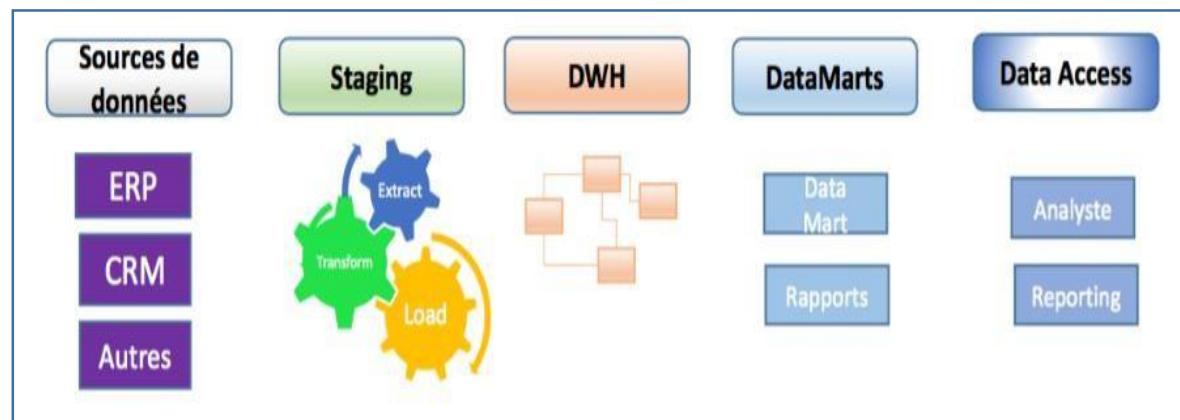
5. Modélisation et conception du Data Warehouse

Lors de la conception d'un Data Warehouse, deux approches se confrontent : la méthode de **Bill Inmon** et celle de **Ralph Kimball**, chacune adaptée à son environnement d'intégration et présentant ses avantages et inconvénients.

5.1 L'approche d'Inmon (ou approche Top-Down)

William H. Inmon (communément appelé Bill Inmon) est un informaticien américain né en 1945, reconnu par beaucoup comme le père du Data Warehouse, ayant contribué à la définition de ses fondements. L'approche d'Inmon, souvent opposée à celle de Ralph Kimball est caractérisée comme étant Top-Down.

Dans son approche, le Data Warehouse est un référentiel centralisé d'entreprise (ou CIF) stockant l'information au niveau le plus détaillé. Des DataMarts modélisés sous forme de schémas en étoile sont ensuite créées à partir de ce Data Warehouse.



Voici la représentation schématique de l'approche :

Figures 7: L'approche d'Inmon (approche Top-Down)

Inmon définit par ailleurs le Data Warehouse via les termes suivants :

« **Subject-Oriented** » : les données du Data Warehouse sont organisées, de telle sorte que tous les éléments se rapportant à un même événement ou objet réel sont liés entre eux ;

« **Time-Variant** » : les différentes modifications apportées aux données de la base de données sont suivies et enregistrées, des rapports synthétiques peuvent ainsi être élaborés ;

« **Non-Volatile** » : les données de la Data Warehouse ne sont jamais ré-écrites ou supprimées une fois engagées, les données sont statiques, en lecture seule et retenues pour les futures rapports de synthèse ;

« **Integrated** » : la base de données contient des données de la plupart ou totalité des applications de l'organisation, ces données œuvrant en parfaite symbiose

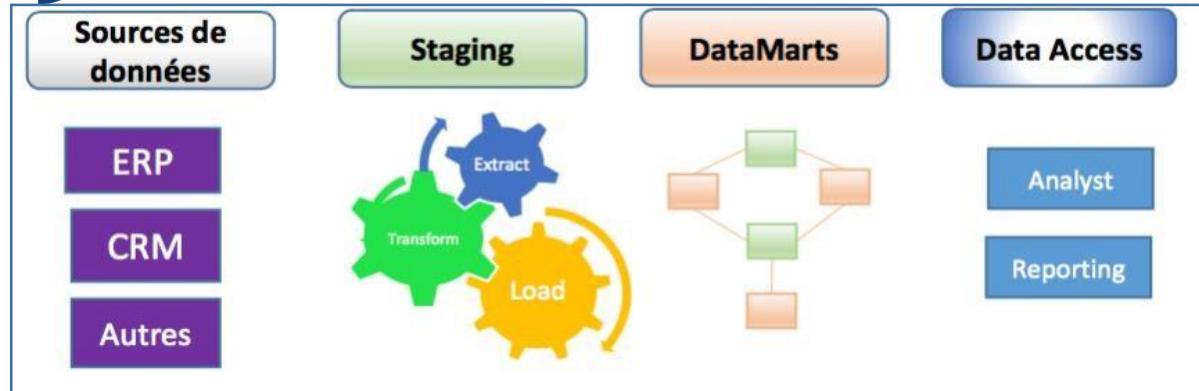
5.2 L'approche de Kimball (ou approche Bottom-Up)

Ralph Kimball est un informaticien et chef d'entreprise américain. Il est connu comme auteur d'ouvrages informatiques, notamment sur le thème du décisionnel.

Son approche s'oppose à celle de William H. Inmon, affirmant qu'un **Data Warehouse doit être rapide et compréhensible**.

Le Data Warehouse peut être vu, selon l'auteur, comme l'union des datamarts cohérents entre eux grâce aux dimensions conformes (*data warehouse bus*). La modélisation dimensionnelle permet un requêtage facile et performant pour les utilisateurs.

Voici la représentation schématique de l'approche :



Figures 8: L'approche de Kimball (ou approche Bottom-Up)

5.3 Inmon vs. Kimball : les caractéristiques majeures

Inmon	Kimball
Commence par la conception du modèle de DW	Commence par la conception du modèle dimensionnel pour les datamarts
Architecture composé d'un staging area permanent, d'un DW et de datamarts dépendants	Architecture qui consiste en un staging area et de datamarts, le DW physique n'existe pas
Le DW est orienté entreprise et les datamarts sont orientés processus	Les datamarts contiennent les données atomiques et agrégées
Le DW contient les données atomiques ; les datamarts les données agrégées	Les datamarts peuvent fournir une vue entreprise ou processus
Le DW utilise un modèle normalisé de toute l'entreprise; les datamarts utilisent des données dimensionnelles orientés sujet	Les datamarts sont implémentés de façon incrémentale et intégrée en utilisant les dimensions conformes
Les utilisateurs peuvent effectuer des requêtes sur le DW et les datamarts	

Figures 9: Inmon vs. Kimball : les caractéristiques majeures

5.4 Avantages inconvenients des 2 approches

	Inmon	Kimball
Construction	Couteux en temps	Rapide
Maintenance	Facile	Difficile redondance à gérer
Coût	Coût initial élevé	Cout initial modéré
Durée mise en œuvre	Long	Court si mode incrémental
Compétences utiles	Equipe spécialisée	Equipe généraliste
Intégration des données	Au niveau entreprise	Par domaine métier

Figures 10: Avantages et inconvenients (approches INMON & KIMBALL)

5.5 Comment bien choisir ?

Malgré une opposition effective des deux modèles, aucun d'entre eux n'est à privilégier et le choix de la meilleure approche dépend de l'activité de l'entreprise ainsi que de ses objectifs à long et court terme sur le plan décisionnel.

L'outil de restitution avec ses possibilités de modélisation peut également influencer le choix d'une des approches. Un modèle hybride est également envisageable (conception du modèle d'entreprise en même temps que les datamarts).

L'expérience a montré que dans le cadre d'un nouveau projet décisionnel, il est nécessaire de mettre en place une solution rapidement. La démarche préconisée dans ce cas est de se focaliser sur la réalisation d'un premier Datamart, soit une approche orientée restitution très proche de celle proposée par Kimball.

Il faut néanmoins voir plus loin et veiller à pouvoir intégrer ce Datamart dans un périmètre élargi fonctionnellement en constituant des dimensions conformes.

5.6 Alimentation du Data Warehouse

Une fois le Data Warehouse conçu, il faut l'alimenter et le charger en données. Cette alimentation (le plus souvent appelée processus ETL « Extract-Transform-Load ») se déroule en 3 phases qui sont :

- Extraction des données primaires (issues par exemple des systèmes de production),
- Transformation des données,
- Le chargement des données traitées dans l'entrepôt de données,

Ces trois étapes décrivent une mécanique cyclique qui a pour but de garantir l'alimentation du Data Warehouse en données homogènes, propres et fiables.

5.6.1 Les phases de l'alimentation « E.T.L. »

Les phases du processus E.T.L. représentent la mécanique d'alimentation du Data Warehouse. Ainsi elles se déroulent comme suit :

5.6.2 L'extraction des données

« L'extraction est la première étape du processus d'apport de données à l'entrepôt de données. Extraire, cela veut dire lire et interpréter les données sources et les copier dans la zone de préparation en vue de manipulations ultérieures. » [Kimball, 2005]. Elle consiste en :

- ✓ Cibler les données,
- ✓ Appliquer les filtres nécessaires,
- ✓ Définir la fréquence de chargement,

Lors du chargement des données, il faut extraire les nouvelles données ainsi que les changements intervenus sur ces données. Pour cela, il existe trois stratégies de capture de changement :

- **Colonnes d'audit** : la colonne d'audit, est une colonne qui enregistre la date d'insertion ou du dernier changement d'un enregistrement. Cette colonne est mise à jour soit par des triggers ou par les applications opérationnelles, d'où la nécessité de vérifier leur fiabilité.
- **Capture des logs** : certains outils ETL utilisent les fichiers logs des systèmes sources afin de détecter les changements (généralement logs du SGBD). En plus de l'absence de cette fonctionnalité sur certains outils ETL du marché, l'effacement des fichiers logs engendre la perte de toute information relative aux transactions.
- **Comparaison avec le dernier chargement** : le processus d'extraction sauvegarde des copies des chargements antérieurs, de manière à procéder à une comparaison lors de chaque nouvelle extraction. Il est impossible de rater un nouvel enregistrement avec cette méthode.

5.6.3 La transformation des données

La transformation est la seconde phase du processus. Cette étape, qui du reste est très importante, assure en réalité plusieurs tâches qui garantissent la fiabilité des données et leurs qualités. Ces tâches sont :

- Consolidation des données.
- Correction des données et élimination de toute ambiguïté.
- Elimination des données redondantes.
- Compléter et renseigner les valeurs manquantes.

Cette opération se solde par la production d'informations dignes d'intérêt pour l'entreprise et de et sont donc prêtes à être entreposées.

5.6.4 Le chargement des données

C'est la dernière phase de l'alimentation d'un entrepôt de données, le chargement est une étape indispensable. Elle reste toute fois très délicate et exige une certaine connaissance des structures du système de gestion de la base de données (tables et index) afin d'optimiser au mieux le processus.

5.7 Politiques de l'alimentation

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

- **Push** : dans cette méthode, la logique de chargement est dans le système de production. Il " pousse " les données vers la zone de préparation quand il en a l'occasion. L'inconvénient est que si le système est occupé, il ne poussera jamais les données.

- **Pull** : contrairement de la méthode précédente, le Pull " tire " les données de la source vers la zone de préparation. L'inconvénient de cette méthode est qu'elle peut surcharger le système s'il est en cours d'utilisation.
- **Push-pull** : c'est la combinaison des deux méthodes. La source prépare les données à envoyer et indique à la zone de préparation qu'elle est prête. La zone de préparation va alors récupérer les données.

5.8 Les outils E.T.L.

Les outils E.T.L, en français E.T.C « Extraction-Transformation-Chargement » [Kimball, 2005], sont des outils qui garantissent la faisabilité et facilitent le déroulement des trois phases citées précédemment. D'où leur importance dans un projet Data Warehouse.

5.9 Mise en œuvre du Data Warehouse

C'est la dernière étape d'un projet Data Warehouse, soit son exploitation. L'exploitation du Data Warehouse se fait par le biais d'un ensemble d'outils analytiques développés autour du Data Warehouse. Donc cette étape nécessite l'achèvement du développement, ou de la mise en place, de ces outils qui peuvent accomplir les fonctions suivantes:

5.9.1 Requêtage ad-hoc :

Le requêtage ad-hoc reste très fréquent dans ce type de projet. En effet, les utilisateurs de l'entrepôt de données, et spécialement les analystes, seront amenés à interagir avec le DW via des requêtes ad-hoc dans le but de faire les analyses requises par leurs métiers et, d'élaborer aussi, des rapports et des tableaux de bord spécifiques. L'accès à ce genre de service peut se faire via différentes méthodes et outils. Cependant, les spécialistes en la matière préconisent de laisser la possibilité à l'utilisateur de choisir les outils qui lui paraissent les plus adéquats.

5.9.2 Reporting :

Destiné essentiellement à la production de rapports et de tableaux de bord, « **il est la présentation périodique de rapports sur les activités et résultats d'une organisation, d'une unité de travail ou du responsable d'une fonction, destinée à en informer ceux chargés de les superviser en interne ou en externe, ou tout simplement concernés par ces activités ou résultants** ».

Ces outils de Reporting ne sont pas, à proprement parler, des instruments d'aide à la décision, mais, lorsqu'ils sont utilisés de manière appropriée, ils peuvent fournir une précieuse vue d'ensemble.

Les rapports sont alors créés par le biais d'outils de Reporting qui permettent de leur donner un format prédéterminé. Les requêtes sont constituées lors de l'élaboration des rapports qui seront ensuite diffusés périodiquement en automatique ou ponctuellement à la demande.

5.9.3 Analyse dimensionnelle des données:

L'analyse dimensionnelle est sans doute celle qui exploite et fait ressortir au mieux les capacités de l'entrepôt de données. Le but par l'analyse dimensionnelle est d'offrir aux utilisateurs la possibilité d'analyser les données selon différents critères afin de confirmer une tendance ou suivre les performances de l'entreprise. Cette analyse se fait selon le principe OLAP, offrant de ce fait aux utilisateurs les possibilités de recourir à différentes opérations facilitant la navigation dans les données. La mise en place de ces outils est une option très intéressante dans la mesure où les données seront accessibles en analyses instantanées. Plusieurs fournisseurs de solution OLAP existent sur le marché et offrent des solutions construites sur des méthodes et technologies différentes. C'est d'ailleurs pour cela que le choix de la solution doit se faire au préalable, selon les besoins en utilisation, la taille de l'entrepôt et les moyens techniques disponibles.

5.9.4 Tableaux de bord :

Les tableaux de bord sont un outil de pilotage qui donne une vision sur l'évolution d'un processus, afin de permettre aux responsables de mettre en place des actions correctives. « Le tableau de bord est un ensemble d'indicateurs peu nombreux conçus pour permettre aux gestionnaires de prendre connaissance de l'état et de l'évolution des systèmes qu'ils pilotent et d'identifier les tendances qui les influenceront sur un horizon cohérent avec la nature de leurs fonctions » [Bouquin, 2003].

Cette forme de restitution a la particularité de se limiter à l'essentiel, c'est-à-dire la mise en évidence de l'état d'un indicateur par rapport à un objectif, tout en adoptant une représentation graphique de l'information.

5.9.5 Data Mining :

Au sens littéral du terme, le Data Mining signifie le forage de données. Le but de ce forage est d'extraire de la matière brute qui, dans notre cas, représente de nouvelles connaissances. L'idée de départ veut qu'il existe dans toute entreprise des connaissances utiles, cachées sous des gisements de données. Le Data Mining permet donc, grâce à un certain nombre de techniques, de découvrir ces connaissances en faisant apparaître des corrélations entre ces données.

Le Data Warehouse constituera alors la première source de données sur laquelle s'exécutera le processus de découverte de connaissances. Dans la majeure partie du temps, l'entrepôt de données représente un pré requis indispensable à toute fouille de données.

Le recours à ce genre de méthode est de plus en plus utilisé dans les entreprises modernes. Les applications et outils implémentant ces solutions sont rarement développés en interne. En effet, les entreprises préfèrent se reposer sur des valeurs sûres du marché afin d'exploiter au plus vite les données en leur possession.

6. Maintenance et expansion

La mise en service du Data Warehouse ne signifie pas la fin du projet, car un projet Data Warehouse nécessite un suivi constant compte tenu des besoins d'optimisation de performance et ou d'expansion. Il est donc nécessaire d'investir dans les domaines suivants [Kimball, 2002]

Support : assurer un support aux utilisateurs pour leur faire apprécier l'utilisation de l'entrepôt de données. En outre, la relation directe avec les utilisateurs permet de détecter les correctifs nécessaires à apporter.

Formation : il est indispensable d'offrir un programme de formation permanent aux utilisateurs de l'entrepôt de données.

Support technique : un entrepôt de données est considéré comme un environnement de production. Naturellement le support technique doit surveiller avec la plus grande vigilance les performances et les tendances en ce qui concerne la charge du système.

Management de l'évolution : il faut toujours s'assurer que l'implémentation répond aux besoins de l'entreprise. Les revues systématiques à certain point de contrôle sont un outil clé pour détecter et définir les possibilités d'amélioration. En plus du suivi et de la maintenance du Data Warehouse, des demandes d'expansion sont envisageables pour de nouveaux besoins, de nouvelles données ou pour des améliorations.

Ces travaux d'expansion sont à prévoir de façon à faciliter l'évolution du schéma du Data Warehouse.

7. Conclusion

On Conclut Le concept « Data Warehouse » est apparu comme une réponse à des besoins grandissants dans le domaine décisionnel. Son adaptabilité et sa capacité de fournir les données nécessaires à une bonne analyse, ont fait de lui un atout majeur et incontournable pour toute entreprise soucieuse du suivi de ces performances.

III. Chapitre III: Analyse et spécifications

Le but de ce chapitre est d'aboutir à un cadrage fonctionnel et technique de notre projet. Ainsi, nous allons d'abord commencer par une description de l'existant en termes de sources de données et de reporting utilisé. Une critique de cet existant permettra de situer les problématiques et limites relativement à ces deux derniers volets. Nous passerons ensuite au recueil des besoins afin de remédier à ces défaillances et de garantir la couverture de l'ensemble des processus positions inclus dans le périmètre de notre projet notamment en matière d'indicateurs et d'axes d'analyse de la solution décisionnelle envisagée. Nous terminerons ce chapitre en précisant une architecture applicative adéquate à notre solution.

1. Étude de l'existant

1.1 Description de l'existant

Dans cette section, nous décrivons l'existant en termes de sources de données ainsi que l'outil de reporting utilisé auparavant.

- ✓ **Sources de données :**

L'étude des différentes sources de données est une phase très importante afin de choisir les tables qui contiennent les données nécessaires à la conception de notre système décisionnel.

Les sources de données sont :

- Les données sont sous formes de 64 Fichier CSV Divisée par Catégories, Société et par années. Chaque Société à 22 fichiers 11 pour chaque catégorie et chaque fichier est pour une année spécifique du 2009 jusqu'au 2019(La société Liander n'a que 20 fichier à cause du manque des donnée de l'année 2009).

- ✓ **Aucun reporting utilisé**

2. Recueil des besoins

Le but de la phase de recueil des besoins est de garantir la couverture fonctionnelle de l'ensemble des processus positions inclus dans le périmètre du projet. Il importe donc de faire un balayage global des besoins définis dans le cahier des charges. En effet, ces besoins ont été complétés et modifiés tout au long des réunions effectuées

2.1 Besoins principaux

- ✓ **Centralisation des sources de données dans un Datawarehouse :**

Puisque les données sont stockées dans différents emplacements, leur consolidation dans un Datawarehouse s'avère primordiale. Ce qui permettra de répondre aux besoins de facilité et de rapidité d'accès aux informations. Ce Datawarehouse constituera une archive riche des données de la direction des ressources humaines.

- ✓ Construction d'une application web contenant les tableaux de bord selon les critères des utilisateurs :

Une fois les données stockées dans le Datawarehouse et les indicateurs sont chargés, les données seront ensuite publiées sous forme de tableaux de bord. Afin d'avoir des informations précises, les utilisateurs demandent la génération de rapports en précisant la distribution des indicateurs et des axes d'analyse sur les tableaux de bord.

- ✓ Facilité d'accès aux données :

Les tableaux de bord seront générés dans le même serveur utilisé par le SIRH. Ainsi, les utilisateurs auront un environnement Simple.

3. Cadrage fonctionnel du projet

En décisionnel, Le but du cadrage fonctionnel est de combler les besoins fonctionnels de notre projet en termes d'indicateurs et d'axes d'analyse. Il s'agit de décrire les indicateurs et axes d'analyse fixés par la Direction relativement au thème énergie.

3.1 Description des indicateurs

Suite aux réunions, nous avons pu fixer les principaux indicateurs ayant un enjeu relatif à la prise de décision.

3.1.1 Indicateurs "énergie"

Les indicateurs relatifs à l'énergie:

- ✓ Indicateurs "Consommation"

Ces indicateurs permettent de mesurer l'évolution de la consommation et d'avoir une vue globale. Le tableau présente les différentes propriétés des deux indicateurs "Consommation".

Indicateurs Propriétés	Consommation	% Consommation pendant les heures à tarif réduit.
Définition	Total Consommation KWh pour l'électricité, m3 pour le gaz	Pourcentage de consommé pendant les heures à tarif réduit. À partir de 22h à 7 h et le week-end.
Enjeu	Mesurer l'évolution de La Consommation	Mesurer l'évolution de La Consommation pendant les heures à tarif réduit.
Périodicité	ANNUAL	
Règle de gestion	Calculer le somme de Consommation des plages de codes postal	Calculer le pourcentage de Consommation à tarif bas des plages de code postal
Source	Fichier CSV : Liander*.csv /Enexis*.csv/Stedin*.csv Champs : annual_consume, annual_consume_lowtarif_perc	
Axes d'analyse	Catégorie, Société, Année, Ville	

Tableau 2 : Indicateurs Consommation

✓ Indicateurs "Connection"

Cette série d'indicateurs, concernant les connections.

Indicateurs Propriétés	nombre de connexions	% de connexions actives
Définition	Total nombre de connexions	pourcentage de connexions actives
Enjeu	Mesurer le nombre des connexions	Mesurer le pourcentage des connexions actives
Périodicité		Annual
Règle de gestion	Calculer la somme de nombre de connexion des plages de codes postal	Calculer le pourcentage de connexion active des plages de code postal
Source	Fichier CSV : Liander*.csv /Enexis*.csv/Stedin*.csv Champs : num_connections, percofactive_connections:	
Axes d'analyse	Catégorie, Société, Année, Ville	

Tableau 3 : *Indicateur Connection*

✓ Indicateurs "Compteurs et livraison"

Cette série d'indicateurs, concernant les compteurs intelligents et livraison d'énergie.

Indicateurs Propriétés	% Compteurs Intelligents	%livraison
Définition	pourcentage des compteurs intelligents	pourcentage de la consommation nette d'électricité ou de gaz. Plus bas, plus d'énergie a été restituée au réseau (par exemple si vous avez des panneaux solaires)
Enjeu	Mesurer le pourcentage des compteurs intelligents	Mesurer le pourcentage d'énergie livré
Périodicité		Annual
Source	Fichier CSV : Liander*.csv /Enexis*.csv/Stedin*.csv Champs : smartmeter_perc, delivery_per:	
Axes d'analyse	Catégorie, Société, Année, Ville	

Tableau 4 : *Indicateur Compteurs et livraison*

3.2 Description des axes d'analyses

Chacun des indicateurs relevés est analysé selon un certain nombre d'axes. Le tableau ci-dessous définit l'ensemble de ces axes d'analyse en précisant pour chacun les différents niveaux d'hiérarchie.

Axe	Description	Niveaux
Temps	La dimension temps	Année
Categorie	Type d'énergie livré	Gaz Electricité
Ville	Nom de la ville	Ville Enexis Liander Stedin
Societe	La Societe Occupe	

Tableau 5 : Axes d'analyses

4. Cadrage technique

4.1 Macro Architecture

Cette architecture respecte la décomposition habituelle de l'architecture constitutive d'un système décisionnel, qui comporte :

- ✓ la phase d'extraction et de transformation des données :

Périodiquement et sur la base des flux de données reçus, des extractions seront réalisées par l'ETL (Extract Transform Load). Ces données extraits subiront par la suite un ensemble de transformations (nettoyage, unification, mise en forme, etc.).

- ✓ la phase de chargement et stockage des données :

Après avoir effectué les traitements nécessaires, les données manipulées seront ensuite chargées grâce à l'ETL pour être stockées dans notre Data Warehouse.

- ✓ La phase de restitution et d'analyse :

C'est à ce point que sont calculés les indicateurs, fournissant ainsi des informations de valeur. Ces informations sont ensuite restituées sous forme de tableaux de bord.

4.2 Architecture Applicative

✓ Alimentation

Pour la préparation de données, nous avons choisi le composant SSIS de l'outil MSBI. En fait, il permet de créer graphiquement des processus de manipulation et de transformation de données puis de générer l'exécutable correspondant. Périodiquement, des scripts ETL sont exécutés afin d'alimenter notre Datawarehouse à partir des sources de données.

✓ Stockage

Les données seront stockées dans le SGBD relationnel SQL selon le mode de stockage ROLAP. Le principe de ROLAP est de conserver toutes les données brutes dans une base de données relationnelles et de calculer les agrégats par des requêtes SQL. ROLAP permet le stockage d'un grand volume de données.

✓ Analyse

Une fois notre Datawarehouse alimenté par l'outil ETL, on peut créer des cubes OLAP qui permettent de faire des analyses multidimensionnelles et dynamiques. Pour la création des cubes, nous allons utiliser le composant SSAS de la suite MSBI aussi.

✓ Restitution

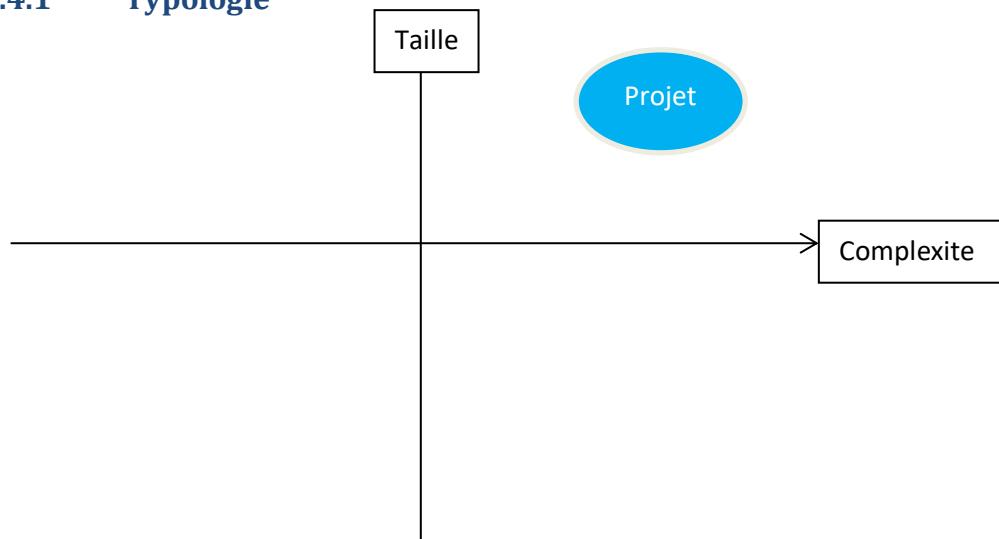
Pour la restitution utilisée dans notre projet est une application web en ASP.net, la bibliothèque ChartsJS et l'outil RStudio. Nous avons choisi ces outils sous la demande de nos clients. En effet La création des tableaux de bord interactifs et très visuels ChartsJS

4.3 Reflexion

C'est grâce à ce projet que nous avons eu l'opportunité de cumuler les connaissances théoriques avec celles de la pratique. Ceci permet également rentrer dans la vie active et de découvrir plus précisément le milieu professionnel.

4.4 Management du projet

4.4.1 Typologie



Figures 11: Typologie du projet

4.4.2 Profile

Petite taille

Faible enjeu

Livrable immatériel

Faible complexité

Faible degré d'innovation

Faible autonomie

Grand taille

Fort enjeu

Livrable matériel

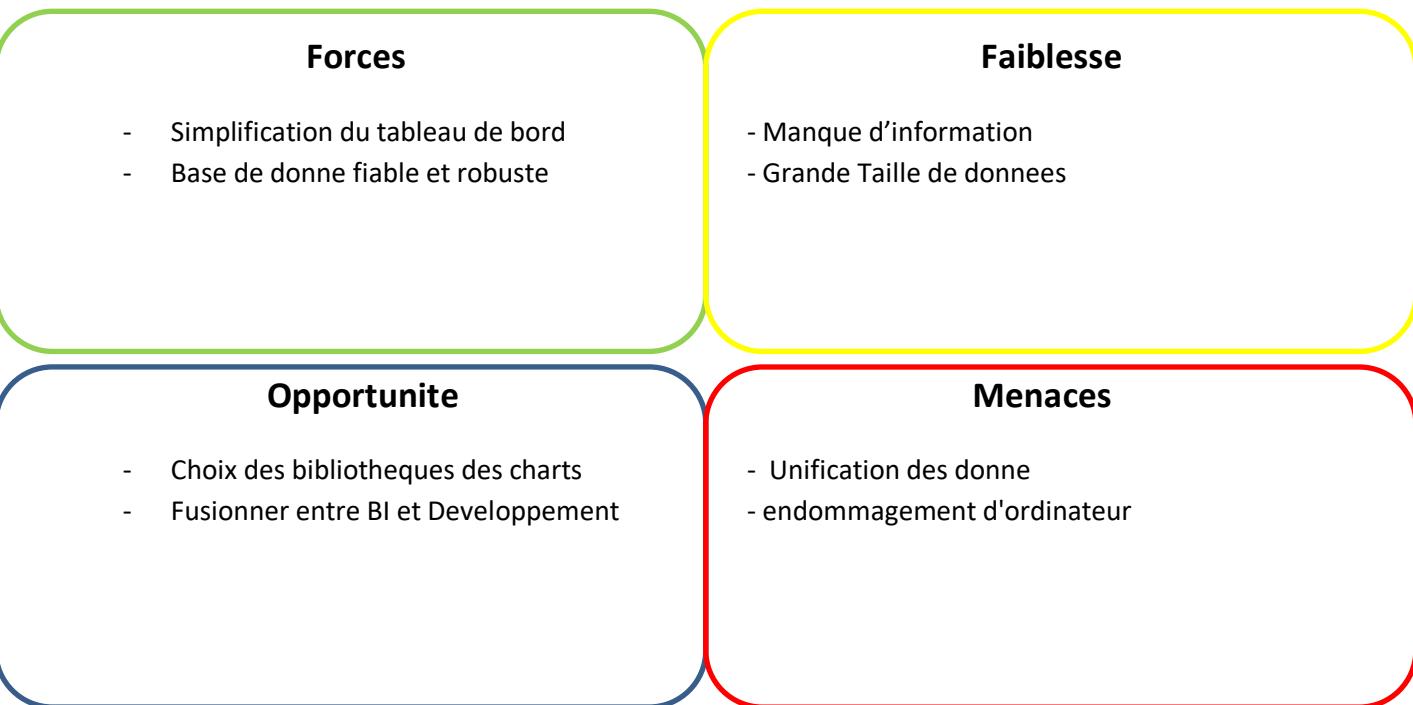
Fort complexité

Fort degré d'innovation

forte autonomie

Figures 12: Profile du projet

4.4.3 Swot



Figures 13: Swot du Projet

5. Conclusion

Dans ce chapitre, nous avons pu voir que la direction des ressources humaines ne dispose pas d'une base de données fiable, contenant toute l'information nécessaire à la production de rapports d'aide à la décision.

Suite à plusieurs réunions, nous avons pu recueillir les besoins en termes de données et de reporting. Nous avons ensuite arrêté en particulier les indicateurs ayant un certain enjeu décisionnel relativement aux thèmes energie, tout en précisant leurs axes d'analyse. Enfin, nous avons défini la macro architecture de notre solution empreinte du cycle dimensionnel et précisé les outils décisionnels adoptés.

Les tâches précisées précédemment constituent l'étape d'analyse et de collecte des besoins, étape cruciale pour le bon déroulement du projet. La définition précise, la collecte fidèle des données lors de ces tâches nous a facilités grandement la phase de conception présentée dans le chapitre suivant.

IV. Chapitre III: Conception

Une fois les besoins définis et analysés, la phase de conception est l'étape suivante dans le cycle dimensionnel. Le but de cette phase est de modéliser les différentes couches du système décisionnel cible. Ainsi, dans ce chapitre, nous présenterons des généralités sur le Datawarehouse et la démarche suivie pour le concevoir. Nous détaillerons notre Datamart conçu. Nous y décrirons enfin la conception de nos tableaux de bord en spécifiant pour chacun les indicateurs et les axes d'analyse mis en jeu.

1. Généralités sur le Datawarehouse

Le Datawarehouse est une collection de données intégrées, orientées sujet, non volatiles, historisées, résumées et disponibles pour l'interrogation et l'analyse.

Il permet de stocker des données nécessaires à la prise de décision. Les données d'un Entrepôt de données respectent donc les caractéristiques suivantes:

-Intégrées: Les données de l'entrepôt proviennent de différentes sources de données et peuvent exister sous différentes formes. Ainsi, il faut les intégrer afin de les homogénéiser.

-Orientées sujet: Les données de l'entrepôt sont structurées par thèmes (les sujets majeurs de l'entreprise).

-Historisées: Le Data Warehouse stocke l'historique des valeurs que la donnée aura prises au cours du temps.

-Non volatiles: Les données de l'entrepôt sont très rarement modifiées, la non volatilité des données est en quelque sorte une conséquence de l'historisation.

-Résumées: Les informations qui proviennent des sources de données doivent être agrégées (ou résumées) afin de faciliter le processus de prise de décision.

- ✓ Approche de conception du Datawarehouse

Les deux approches les plus connues en conception de Data Warehouse sont :

- L'approche basée sur les besoins d'analyse
- L'approche basée sur les sources des données

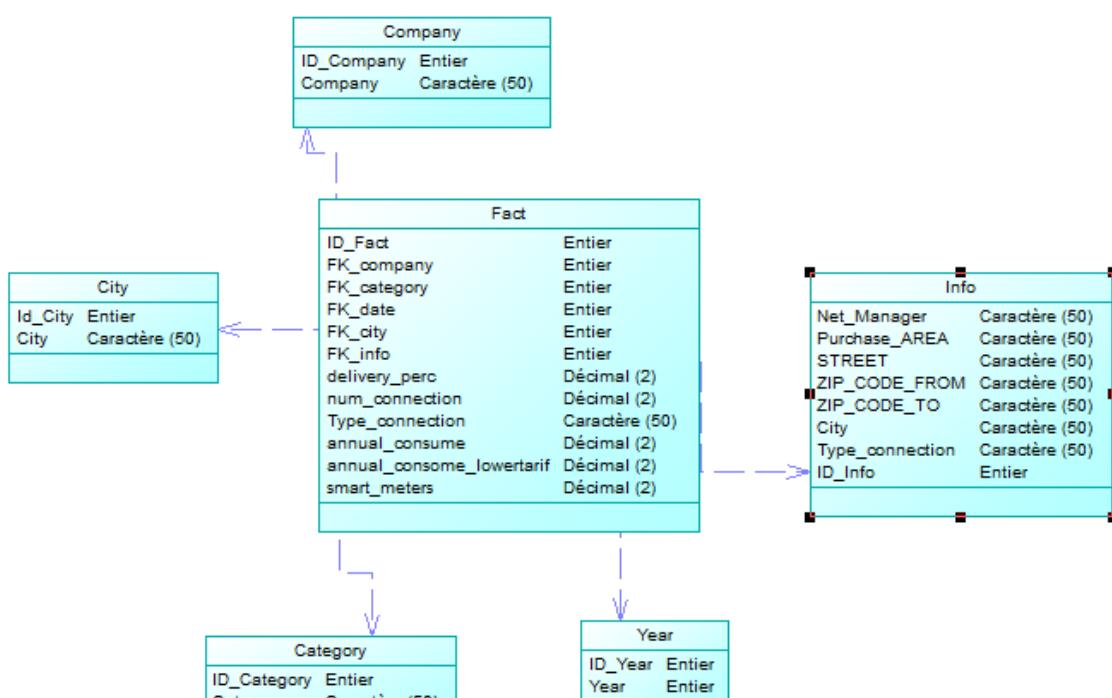
L'approche la plus convenable à notre projet est celle basée sur les besoins d'analyse. Elle est aussi appelée « approche descendante » (*Top-Down Approach*) et est proposée par Ralph KIMBALL.

Pour Ralph KIMBALL, le data warehouse est constitué de l'agrégation des différents Datamarts. Les Datamarts sont les subdivisions logiques du Data Warehouse. KIMBALL met en avant le développement itératif et incrémental du Data Warehouse, chaque itération devant correspondre au traitement d'un sujet particulier et donc à la création d'un nouveau Datamart.

Les datamarts sont placés au centre de l'architecture. Le reste sera composé d'un *staging area* temporaire. Ce dernier est la zone où la plupart des opérations de nettoyage et de préparation des données ont lieu avant leur chargement dans le Datawarehouse. Le staging area est dit temporaire car les données sont détruites une fois le chargement des datamarts terminé.

2. Le modèle en étoile:

Le modèle en étoile est une représentation fortement dénormalisée qui assure un haut niveau de performance des requêtes même sur de gros volumes de données.



Figures 14: Modèle en étoile

3. Modélisation du Datawarehouse

3.1 Matrice d'expression des besoins

A partir des besoins répertoriés dans le deuxième chapitre, nous avons lié chaque indicateur à un certain nombre d'axes d'analyse recensés. Le *tableau 5* présente la matrice des liaisons relevées. Cette matrice permet de faciliter la conception du Datawarehouse

Axes d'analyse	Temps	Catégorie	Ville	Société
Indicateurs				
Consommation	X	X	X	X
% Consommation pendant les heures à tarif réduit	X	X	X	X
% de connexions actives	X	X	X	X
nombre de connexions	X	X	X	X
% Compteurs Intelligents	X	X	X	X
%livraison	X	X	X	X

Tableau 6 : *Matrice d'expression des besoins*

3.2 Schéma dimensionnel

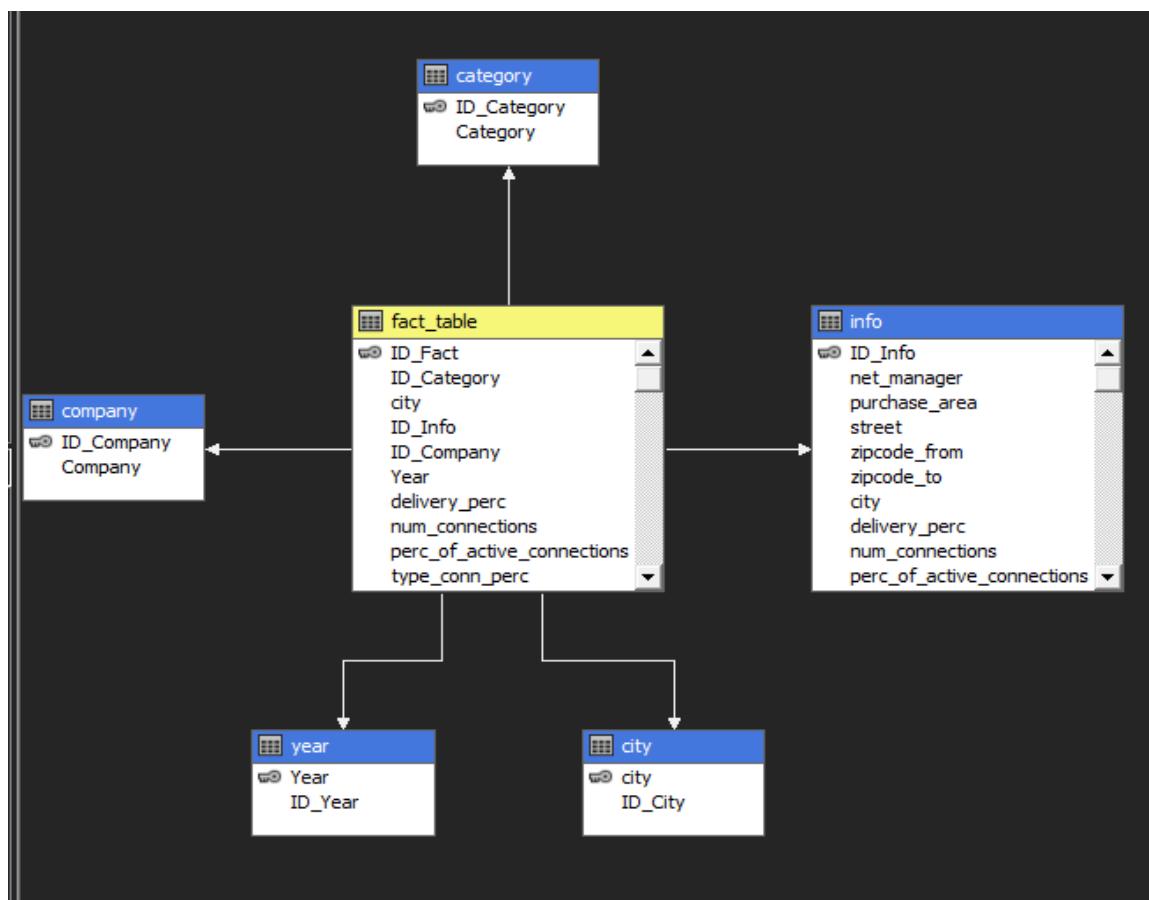
Un datamart, ou magasin de données, permet de stocker toute l'information utile, relative à un thème précis. Avant d'être intégrée dans le magasin de données, l'information est tout d'abord extraite des bases sources et nettoyée. Puis, elle est mise en forme de manière à l'adapter aux besoins de l'utilisateur final. Cette mise en forme passe par la modélisation des données en schémas dimensionnels plus proches des utilisateurs finaux et adaptés à leurs besoins.

Pour cela, nous passons par une modélisation des données en un schéma dimensionnel, car elle est la plus appropriée aux requêtes et analyses effectuées sur les entrepôts de données. Elle est simple à

créer, stable et intuitivement compréhensible par les utilisateurs finaux. Un schéma dimensionnel consiste en une ou plusieurs « tables de faits » qui contiennent les faits ou les mesures à calculer. Ces mesures serviront plus tard à calculer les indicateurs dont nous avons besoin. Ces tables de faits sont entourées par d'autres tables, appelées « dimensions », qui contiennent les éléments descriptifs des faits. En effet, le modèle ressemble à une étoile.

3.3 Modélisation des datamarts

Toutes les mesures de nos tables de faits sont calculées par rapport aux axes d'analyse recensés dans la phase d'étude des besoins et qui sont modélisés sous forme de tables de dimension. Pour chaque table de dimension, nous avons créé une clé primaire, appelée clé technique, différente de celle de la table source, appelée clé fonctionnelle. La clé technique est une séquence qui s'auto-incrémente et sert à indexer les enregistrements de la dimension et à améliorer la performance de la recherche. Nous avons toutefois gardé la clé fonctionnelle pour permettre les futures jointures. On va présenter dans ce qui suit le Datamart conçus dans notre projet



Figures 15: Datamart

4. Conception des tableaux de bord

Le modèle du Datamart établi ne suffit pas pour clôturer la phase de conception et passer à la réalisation. En effet, il nous faut décider de la meilleure manière de présenter les indicateurs relevés

à l'utilisateur final. Cette présentation doit se faire de la façon la plus simple et la plus expressive possible Ainsi, nous avons conçu plusieurs tableaux de bord dont chacun a un objectif précis.

Les tableaux ci-dessous résument la conception des tableaux de bord à réaliser. Pour chaque tableau de bord, nous avons indiqué l'indicateur et les axes d'analyse concernés. Nous avons également explicité l'objectif et les résultats attendus de chacun.

3.1 Tableaux de bord 'Accueil'

Le tableau ci-dessous présente le tableau de bord d'accueil.

N° du tableau de bord	Intitulé du tableau de bord	Indicateurs	Axes d'analyse	Objectifs et résultats attendus
1	Tableaux de bord d'accueil	Consommation	Temps Catégorie Société Ville	Suivre la tendance de la consommation par année et évolution de la Consommation Par société
		Compteurs et livraison Connection	Catégorie Temps Catégorie	Connaitre Pourcentage de livraison et Pourcentage d'utilisation des compteurs intelligents dans la dernière année et pour les deux catégories Connaitre le nombre de connexion intelligent dans la dernière année et pour les deux catégories

Tableau 7 : Tableaux de bord "Accueil"

3.2 Tableaux de bord "Société"

Le tableau 7 liste les tableaux de bord des sociétés, en présentant les indicateurs traités par chacun et les axes d'analyses utilisés.

N° du tableau de bord	Intitulé du tableau de bord	Indicateurs	Axes d'analyse	Objectifs et résultats attendus
1	Tableaux de bord des sociétés	Consommation	Temps Catégorie Ville	Suivre la tendance de la consommation par année et Connaitre les villes les plus consommateurs

Tableau 8 : *Tableaux de Bord des 'Sociétés'*

3.3 Tableaux de bord "2T"

Le tableau 6 présente les tableaux de bord traitant la partie Employée (Production)

N° du tableau de bord	Intitulé du tableau de bord	Indicateurs	Axes d'analyse	Objectifs et résultats attendus
3	Tableaux de bord des 2T	Consommation Compteurs et livraison Connection	Temps Catégorie Ville	Suivre la tendance de la

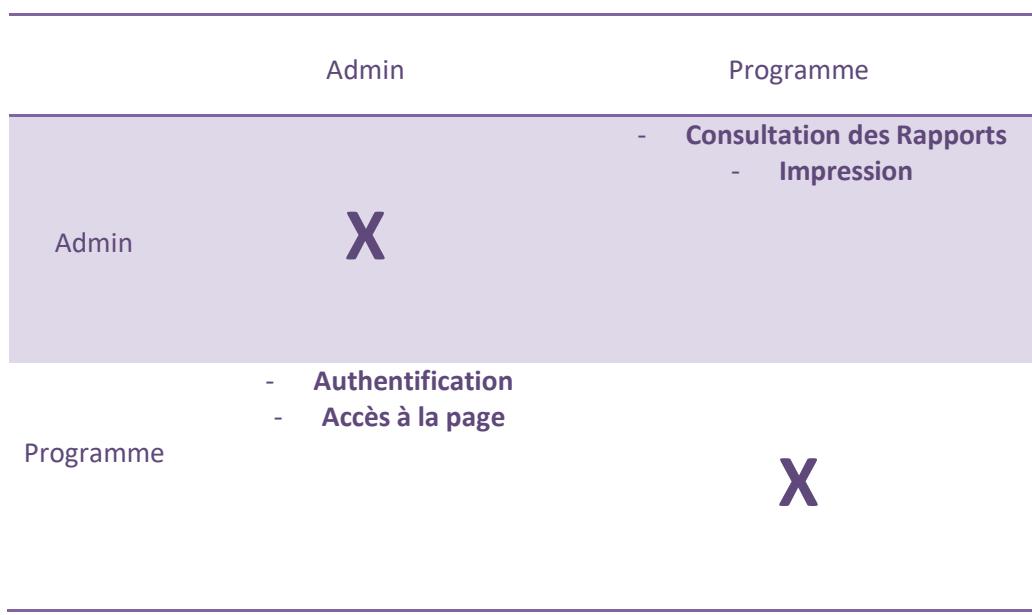
Tableau 9 : *Tableaux de bord "2T"*

3.4 Conception d'application WEB

3.4.1 Introduction

La conception de logiciel met en œuvre un ensemble d'activités qui à partir d'une demande d'informatisation d'un processus (demande qui peut aller de la simple question orale jusqu'au cahier des charges complet) permettent la conception, l'écriture et la mise au point d'un logiciel (et donc de programmes informatiques) jusqu'à sa livraison au demandeur.

3.4.2 Choix de la méthodologie de conception



Figures 16: Matrice des flux

N° Fonction	Fonctionnalité	Détail	Acteur
Projet de fin d'étude	{ 60 }		2019/2020

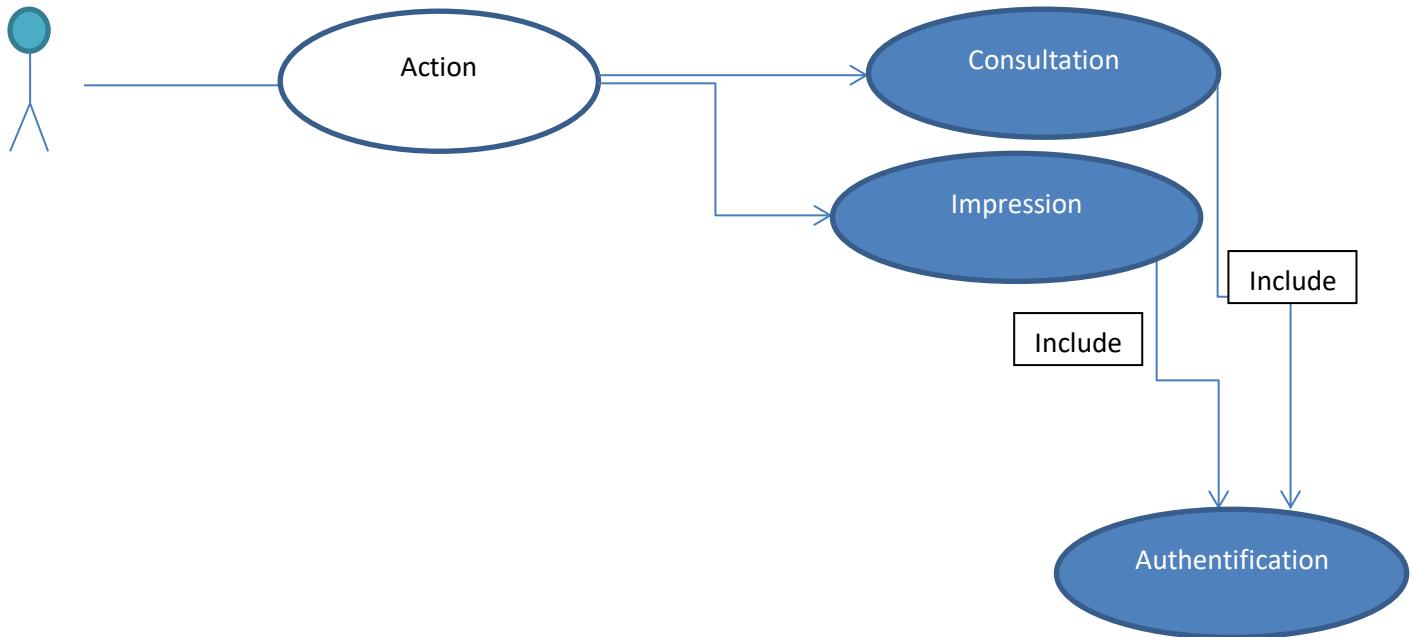
F1	Authentification	L'administrateur doit taper son Identificateur et Mot de Passe	Admin
F2	Ouverture du site web	Le programme donne l'accès au site	Programme
F3	Consultation	L'admin peut consulter les rapports	Admin
F4	Impression	L'admin peut Imprimer les rapport	Admin

Figures 17: Tableau des fonctions automatisées

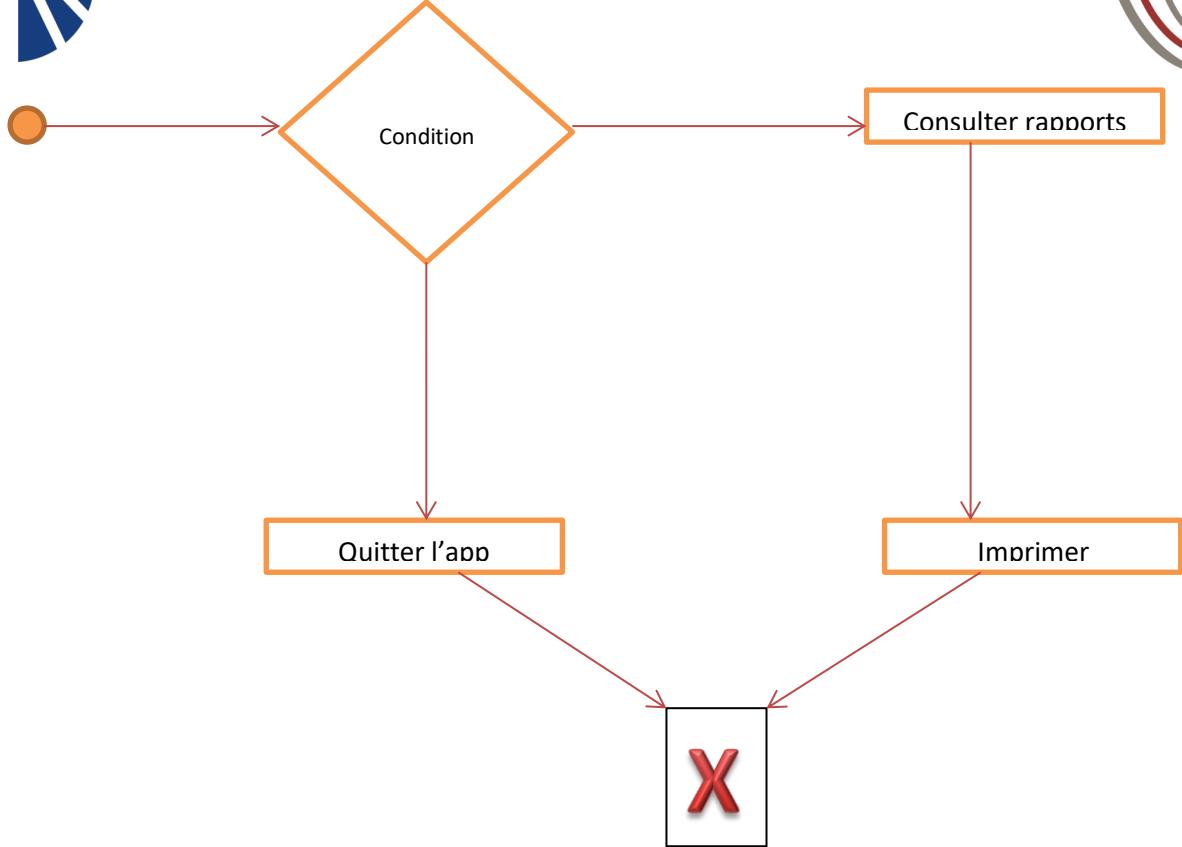
Sigle	Libelle	Type
net_manager	code du gestionnaire de réseau régional	Varchar
purchase_area	code de la zone où l'énergie est achetée	Varchar
street	Name of the street	Varchar
zipcodefrom	la gamme de codes postaux couverts, 4 chiffres et 2 lettres	Varchar
zipcodeto	la gamme de codes postaux couverts, 4 chiffres et 2 lettres	Varchar
city	Name of the city	Varchar
num_connections	Nombre de connexions dans la plage de codes postaux	INT
delivery_perc	pourcentage de la consommation nette d'électricité ou de gaz. Plus bas, plus d'énergie a été restituée au réseau	Double
percofactive_connections	Pourcentage de connexions actives dans la plage de codes postaux	Double

typeofconnection	principal type de connexion dans la plage de codes postaux	Varchar
typeconnperc	pourcentage de présence du principal type de connexion dans la plage de codes postaux	Varchar
annual_consume	Consommation annuelle. Kwh pour l'électricité, m3 pour le gaz	Double
annualconsumelowtarif_perc	Pourcentage de consommation pendant les heures à tarif réduit. À partir de 22h à 7 h et le week-end.	Double
smartmeter_perc	pourcentage de compteurs intelligents dans les plages de codes postaux	Double

Figures 18: Dictionnaires de données



Figures 19: Diagramme de cas d'utilisation



Figures 20: Diagrammes d'état transition

5. Conclusion

Grâce aux indicateurs et axes d'analyse identifiés dans le chapitre précédent, nous avons pu ressortir la matrice d'expression des besoins et modéliser par la suite notre Datawarehouse selon la démarche descendante de Kimball. Nous avons détaillé notre Datamart conçus. Nous avons également abordé la conception de différents tableaux de bord relativement au thème énergie en précisant l'objectif, les indicateurs et les axes d'analyse pour chacun d'entre eux. Après cette phase de conception, nous passons à la phase de mise en œuvre, exposée en détail dans le chapitre suivant.

V. Chapitre IV: Mise en oeuvre

Dans ce chapitre, nous présentons la phase de la réalisation qui concrétise la conception exposée lors du chapitre précédent. Nous décrivons la mise en œuvre du datawarehouse, et d'application web en explicitant les étapes suivies pour notre processus ETL. Nous illustrons également la réalisation de nos cubes OLAP et la construction de notre siteweb.

1. Réalisation de l'ETL

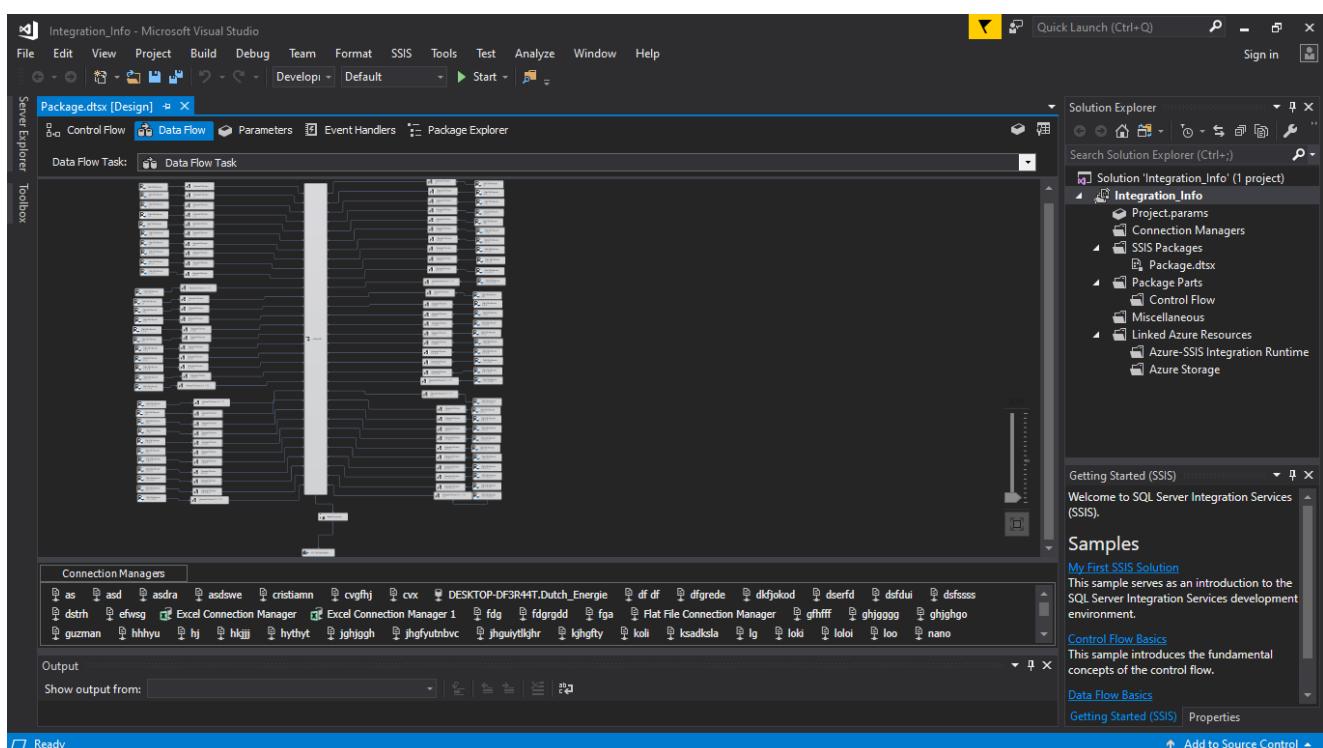
Comme déjà expliqué, l'ETL est le processus qui nous permet d'extraire les données depuis les sources, les transformer et les charger dans le datawarehouse. Avant de s'y attaquer, nous avons créé la base de données sur le serveur de test dédié. L'ETL est un processus qui sera lancé plusieurs fois, de façon périodique, pour que les nouvelles données issues des sources soient incluses dans notre datawarehouse.

Pour notre réalisation, nous avons procédé par étapes. En effet, nous avons alimenté d'abord les tables de dimension, puis la table de faits. Ces étapes sont détaillées dans ce qui suit.

1.1 Alimentation des tables de dimension

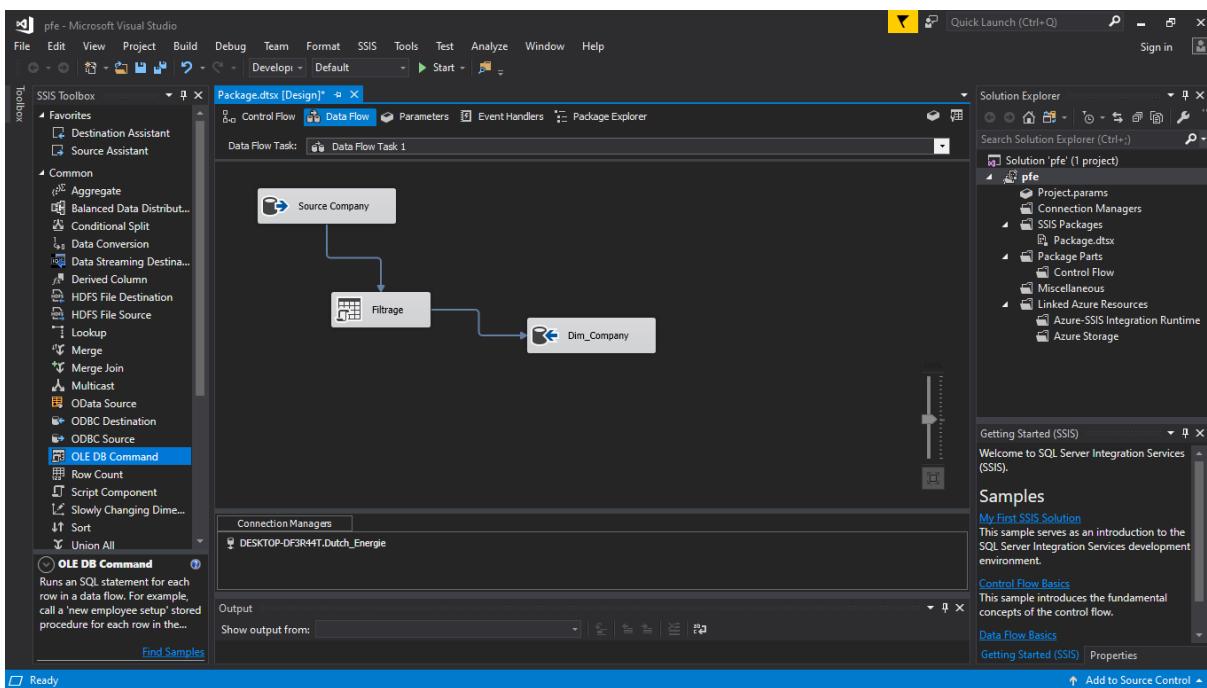
Il s'agit de remplir nos tables de dimensions à partir des sources de données. Dans cette étape, nous effectuons en même temps et les transformations nécessaires à nos dimensions et leur chargement. Pour le rafraîchissement des données de ces tables, tout en insérant que les données non existantes dans les tables de dimensions,

Mais avant de commencer ce processus, nous devions convertir les données des fichiers CSV vers SQL SERVER pour faciliter la manipulation des données et gagner du temps



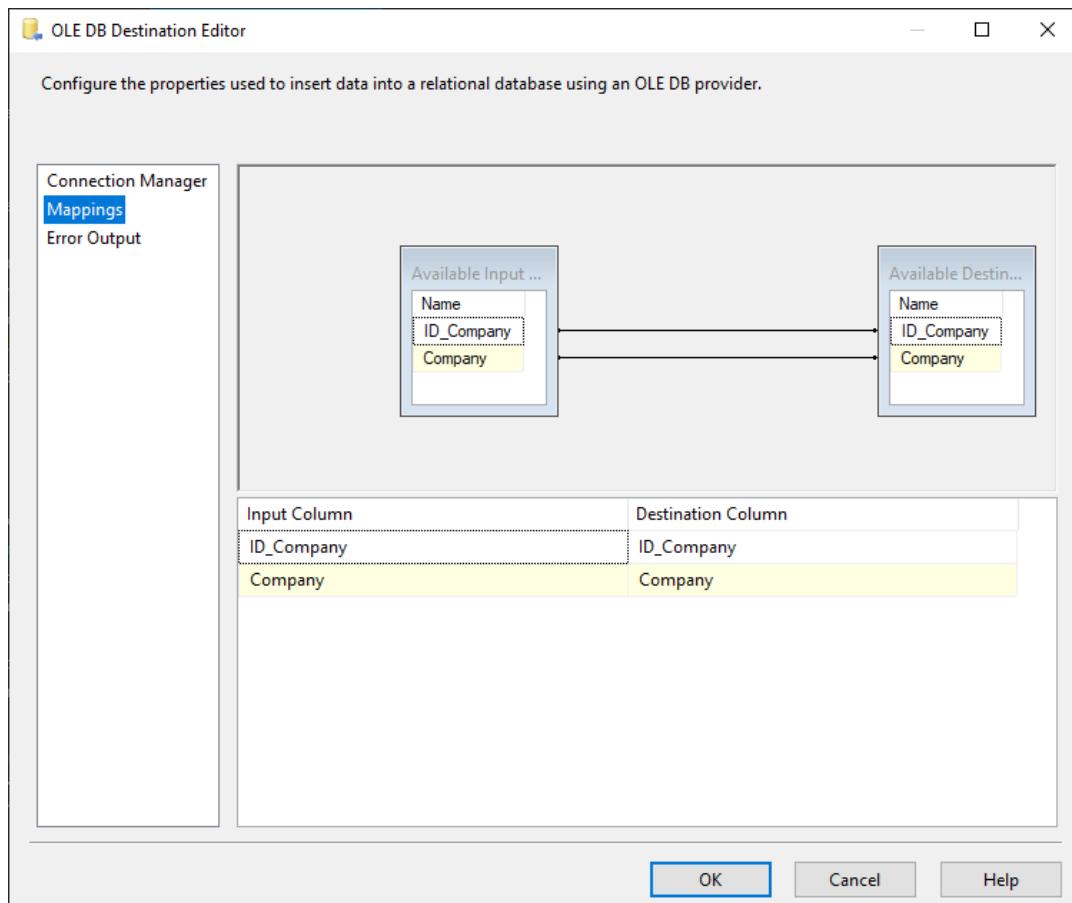
Figures 21: Conversion des données Csv Vers SQL Server

Prenons l'exemple d'alimentation de la table de dimension Company illustré dans la figure 5



Figures 22: Table Dimension Comapany

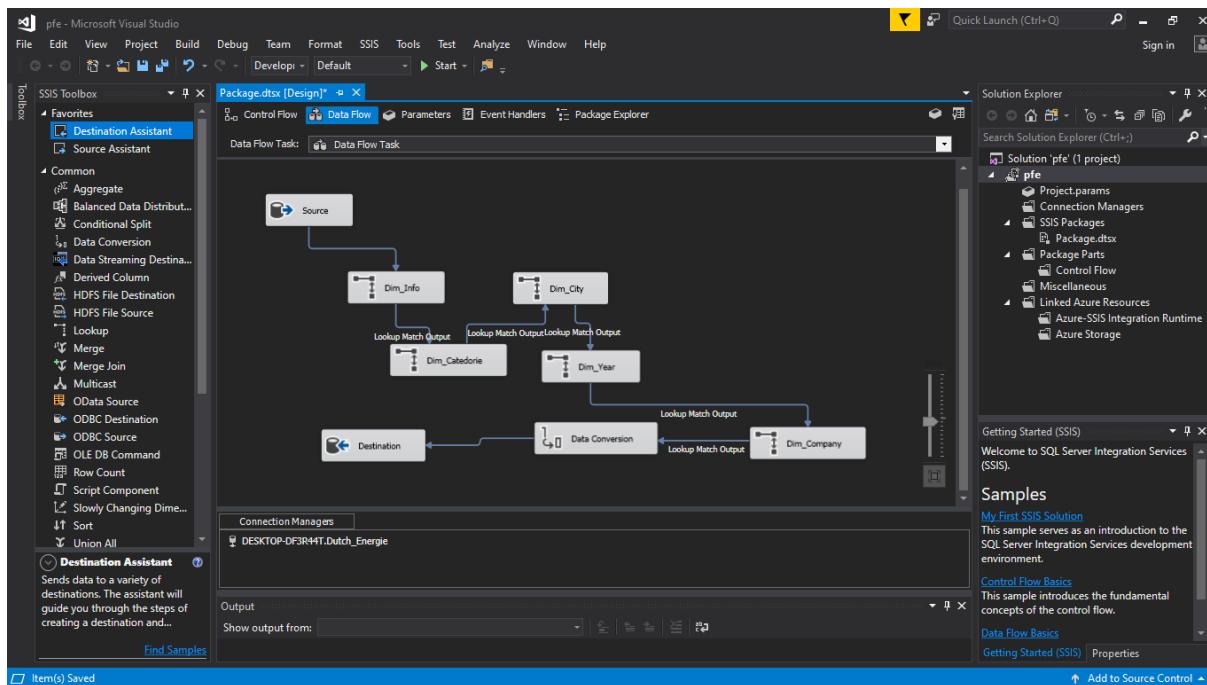
Pour alimenter la dimension Company, on commence par récupérer lesdonnées company à partir de l'ancienne base de données. Pour ce faire, on utilise le composant Source OLE DB. Puis, on utilise le composant Commande OLE DB pour supprimer les doublons et pour lier les données en entrées avec les données que l'on veut avoir en sortie en utilisé destination OLE DB voilà comme illustré dans la figure 6.



Figures 23: Mapping Table Dimension Company

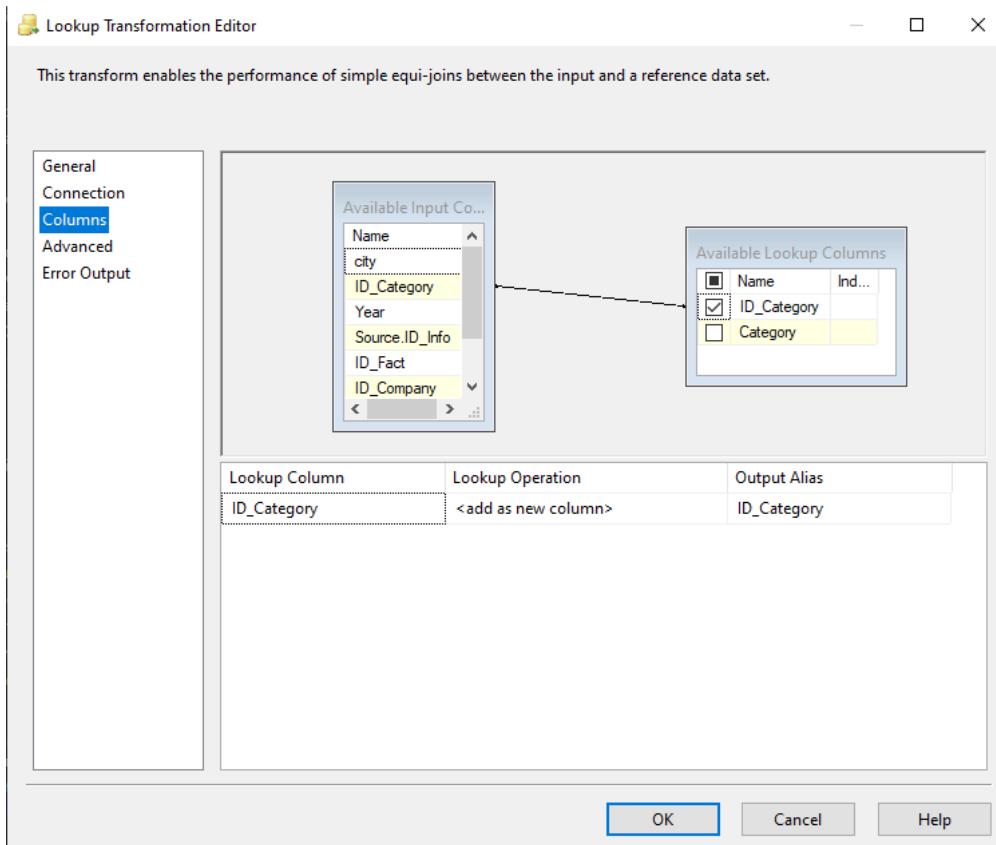
1.2 Alimentation de table de faits

L'alimentation des tables on a suivi la même démarche pour alimenter toutes les tables de faits. Pour ce faire, on commence par alimenter les tables intermédiaires. Ces dernières servent à stocker toutes les données nécessaires à l'alimentation des tables de faits. Prenons comme exemple la table de faits . La figure ci-dessous l'illustre.



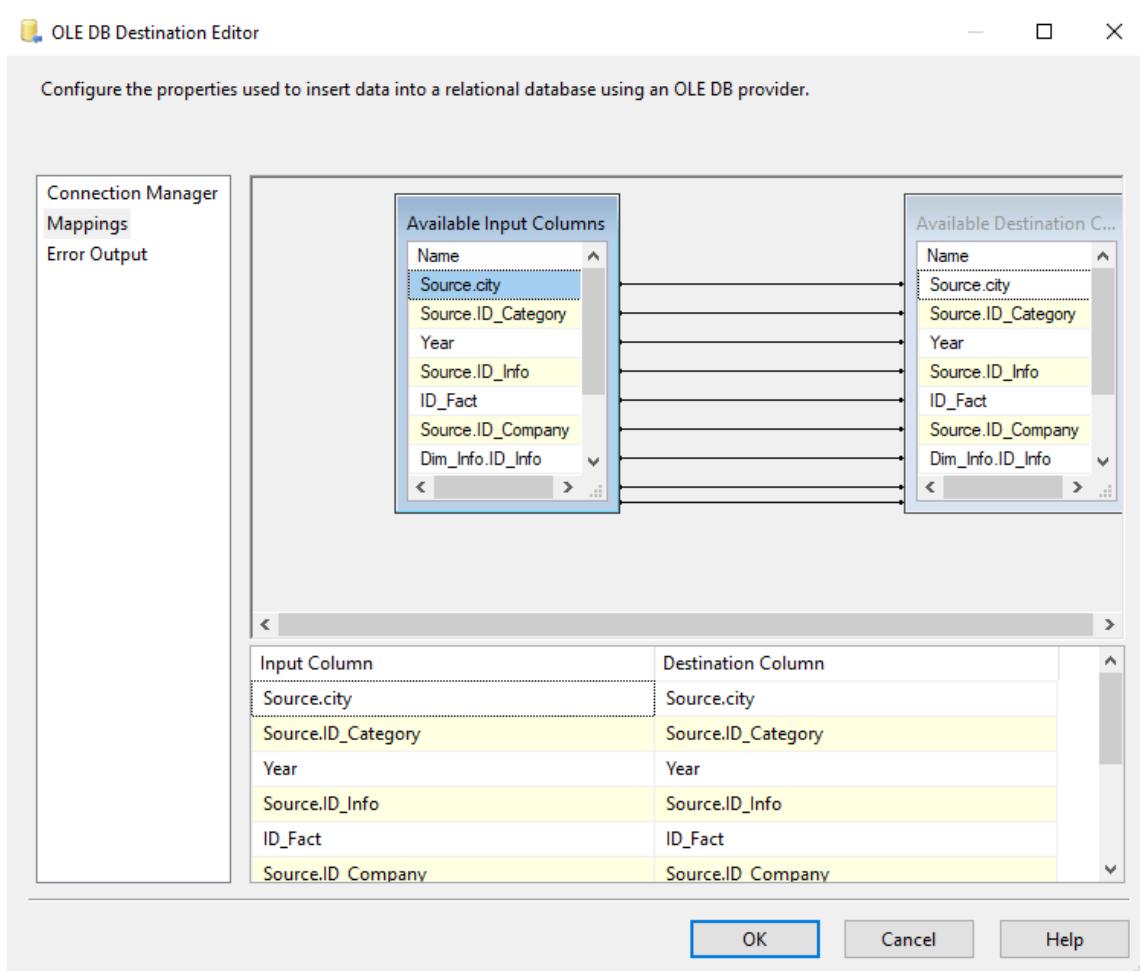
Figures 24: Table de fait

Pour alimenter la table de fait, on commence par récupérer les données à partir des dimensions liées. Pour ce faire, on utilise le composant Source OLE DB. Puis, on utilise le composant Recherche appelé aussi (lookup) pour récupère les mesure et les identificateurs comme illustré dans la figure 25.



Figures 25: Lookup Categorie

On utilise ensuite Destination OLE DB pour lier les données en entrées avec les données que l'on veut avoir en sortie.



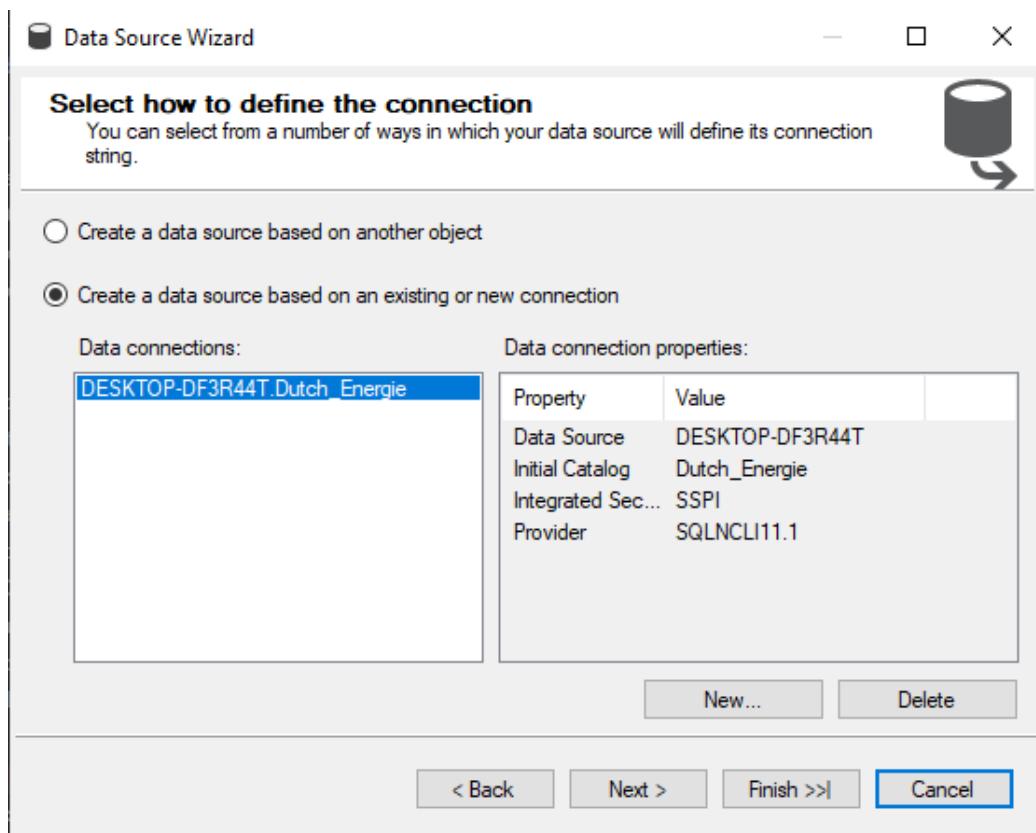
Figures 26: Mapping table de fait

2. Cubes OLAP

Dans SSAS, un cube OLAP est un rapport dynamique basé sur les croisements des axes. Il permet de naviguer au sein de grands volumes de données des plus agrégées au plus détaillées.

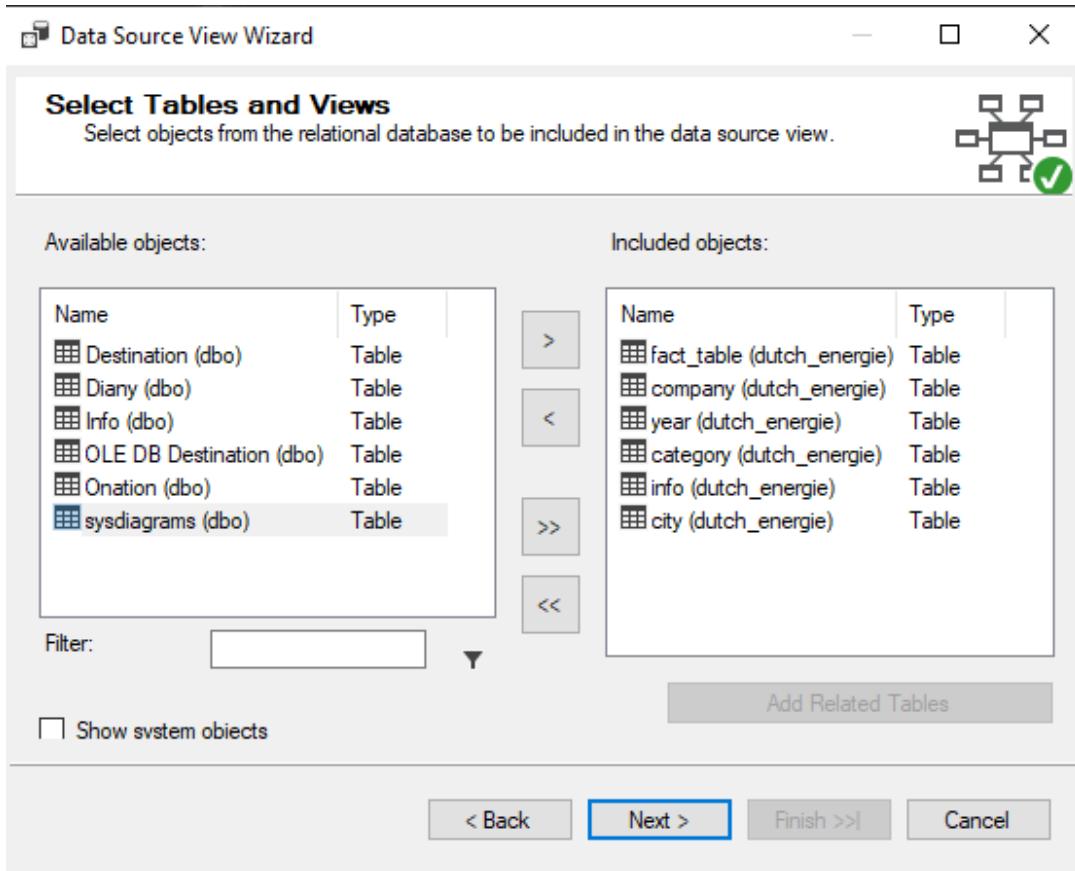
2.1 Crédit des cubes OLAP

Tout d'abord, nous commençons par configurer la connexion à la base de données.



Figures 27: Connexion à la base de données

Ensuite, on crée une vue ou bien schéma de source de donnée



Figures 28: Schéma de source de donnée

Pour créer un cube, on doit d'abord indiquer sa table de faits et les dimensions partagées qu'il contient. On doit également créer ses propres dimensions, c.-à-d. les dimensions non existantes dans les autres cubes. Déployer et Enfin, il ne nous reste plus qu'à insérer la mesure que nous souhaitons calculer.

Cube Wizard

Completing the Wizard

Name the cube, review its structure, and then click Finish to save the cube.



Cube name:

Dutch Energie 1

Preview:

- Measure groups
 - Fact Table
 - Delivery Perc
 - Num Connections
 - Perc Of Active Connections
 - Annual Consume
 - Annual Consume Lowtarif Perc
 - Smartmeter Perc
- Dimensions
 - Year
 - City
 - Company
 - Category

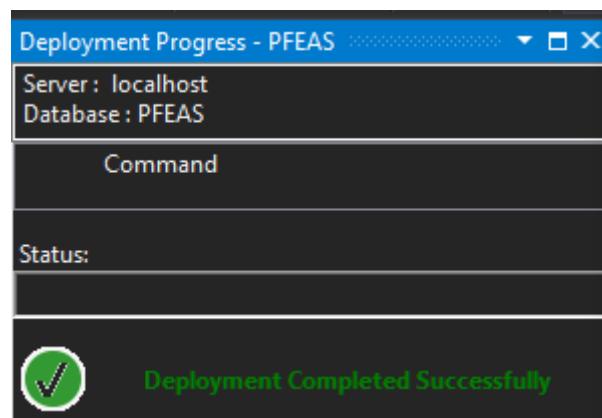
< Back

Next >

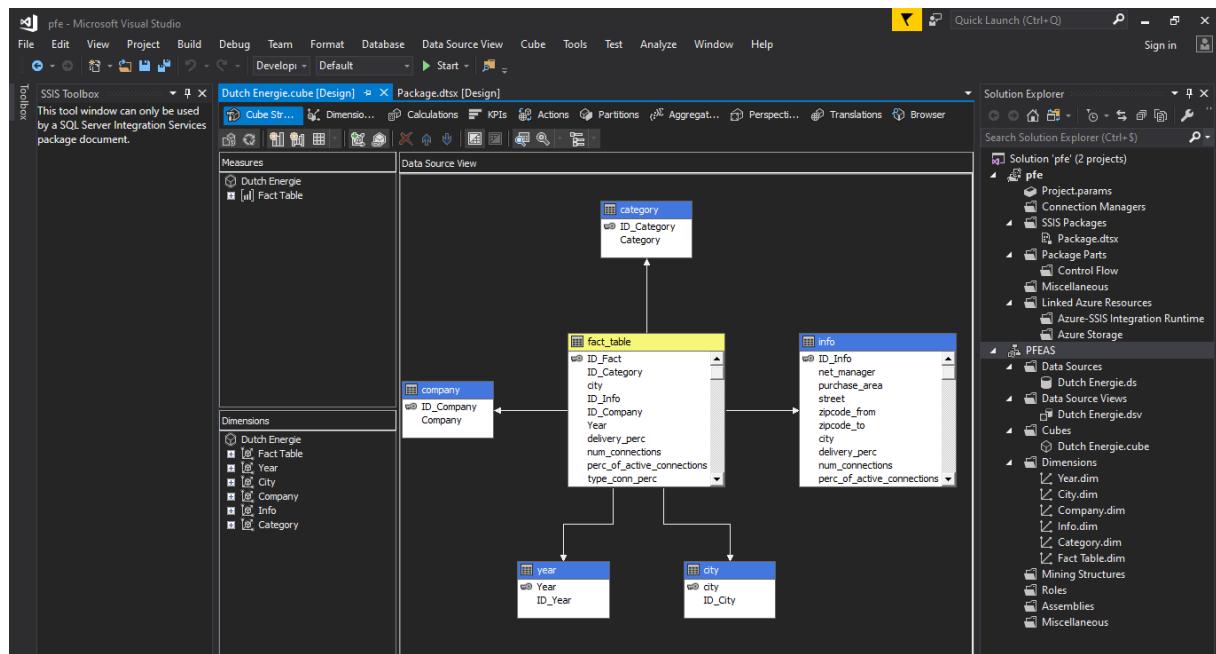
Finish

Cancel

Figures 29: Creation de cube



Figures 30: Deployment



Figures 31: Figure 1: Cube

3. Rstudio

R est un langage de programmation et un environnement logiciel gratuit pour le calcul statistique et les graphiques pris en charge par la Fondation R pour le calcul statistique.

3.1 Ggplot2

Pour visualisez les données on a utilisé Ggplot2. R a plusieurs systèmes pour créer des graphiques, mais ggplot2 est l'un des plus élégants et des plus polyvalents. Ggplot2 implémente la grammaire des graphiques, un système cohérent pour décrire et construire des graphiques. Avec ggplot2, vous pouvez faire plus rapidement en apprenant un système et en l'appliquant à de nombreux endroits.

3.2 KableExtra

Le but de kableExtra est de vous aider à créer des tableaux complexes communs et à manipuler les styles de tableau. Il importe le symbole pipe%>% de magrittr et verbalise toutes les fonctions, donc en gros vous pouvez ajouter des "couches" à une sortie kable d'une manière similaire à ggplot2 et plotly

```

1 --
2 title: 'PFE'
3 output: html_document
4 runtime: shiny
5
6 ---
7 #Load Libraries
8
9 ```{r ,message=FALSE,warning=FALSE}
10
11 library(tidyverse)
12 library(knitr)
13 library(kableExtra)
14
15
16 fillcolor = "#1B9E77"
17 fillcolor2 = "#fff65a"
18
19 ```
20

```

Figures 32: Chargement des bibliothèque sur RSTUDIO

```

enexis_electricity_01012019 <- read_csv("../dutch_energy/Electricity/enexis_electricity_01012019.csv")
enexis_electricity_01012019$year = 2019

liander_electricity_01012019 <- read_csv("../dutch_energy/Electricity/liander_electricity_01012019.csv")
liander_electricity_01012019$year = 2019

stedin_electricity_01012019 <- read_csv("../dutch_energy/Electricity/stedin_electricity_2019.csv")
stedin_electricity_01012019$year = 2019

electricity_2019 <- rbind(enexis_electricity_01012019,liander_electricity_01012019,stedin_electricity_01012019)

glimpse(electricity_2019)
...

```

Tableau 10 : *Chargement des données sur RSTUDIO*

```
electricity_2019 %>%
group_by(city) %>%
summarise(cityconsumption = sum(annual_consum) / 1e6) %>%
arrange(desc(cityconsumption)) %>%
mutate(city = reorder(city, cityconsumption)) %>%
head(10) %>%
kable() %>%
kable_styling(full_width = T) %>%
row_spec(1, bold = T, color = "white", background = "#1B9E77")
```

Tableau Html en utilisant
KableEXTRA

```
ts %>%
group_by(city) %>%
summarise(cityconsumption = sum(annual_consum) / 1e6) %>%
arrange(desc(cityconsumption)) %>%
mutate(city = reorder(city, cityconsumption)) %>%
head(10) %>%

ggplot(aes(x = city, y = CityConsumption)) +
geom_bar(stat = 'identity', colour = "white", fill = #1B9E77) +
geom_text(aes(x = city, y = 1, label = paste0("(", round(cityconsumption, 2), " )", sep = "")),  

  hjust = 0, vjust = .5, size = 4, colour = 'white',
  fontface = 'bold') +
labs(x = 'city',
  y = 'city',
  title = 'city and Annualconsumption',
  subtitle = "Consumption in millions of kwh") +
coord_flip() +
theme_bw()
```

Visualisation en
utilisant Ggplot

Tableau 11 : *Traçage est visualisation*

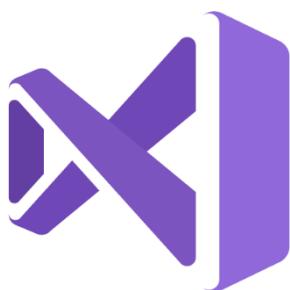
4. Tableaux de Bord

3.1 Crédation des rapports

Les rapports statiques sont des rapports clairs et précis. Ils sont destinés à des utilisateurs qui n'ont pas besoin de faire des analyses en naviguant dans le cube. La création de ces rapports est faite via RStudio et Chartjs.

3.2 Outils de Travail

Visual studio



Microsoft Visual Studio est une suite de logiciels de développement pour Windows et mac OS conçue par Microsoft. La version que j'ai utilisée s'appelle Visual Studio 2019.

Rstudio



RStudio est un environnement de développement gratuit, libre et multiplateforme pour R, un langage de programmation utilisé pour le traitement de données et l'analyse statistique. Il est disponible sous la licence libre AGPLv3, ou bien sous une licence commerciale, soumise à un abonnement annuel.

ChartJS



Chart.js est une bibliothèque JavaScript open source gratuite pour la visualisation des données, qui prend en charge 8 types de graphiques: barre, ligne, zone, tarte (beignet), bulle, radar, polaire et scatter.

ASP.net



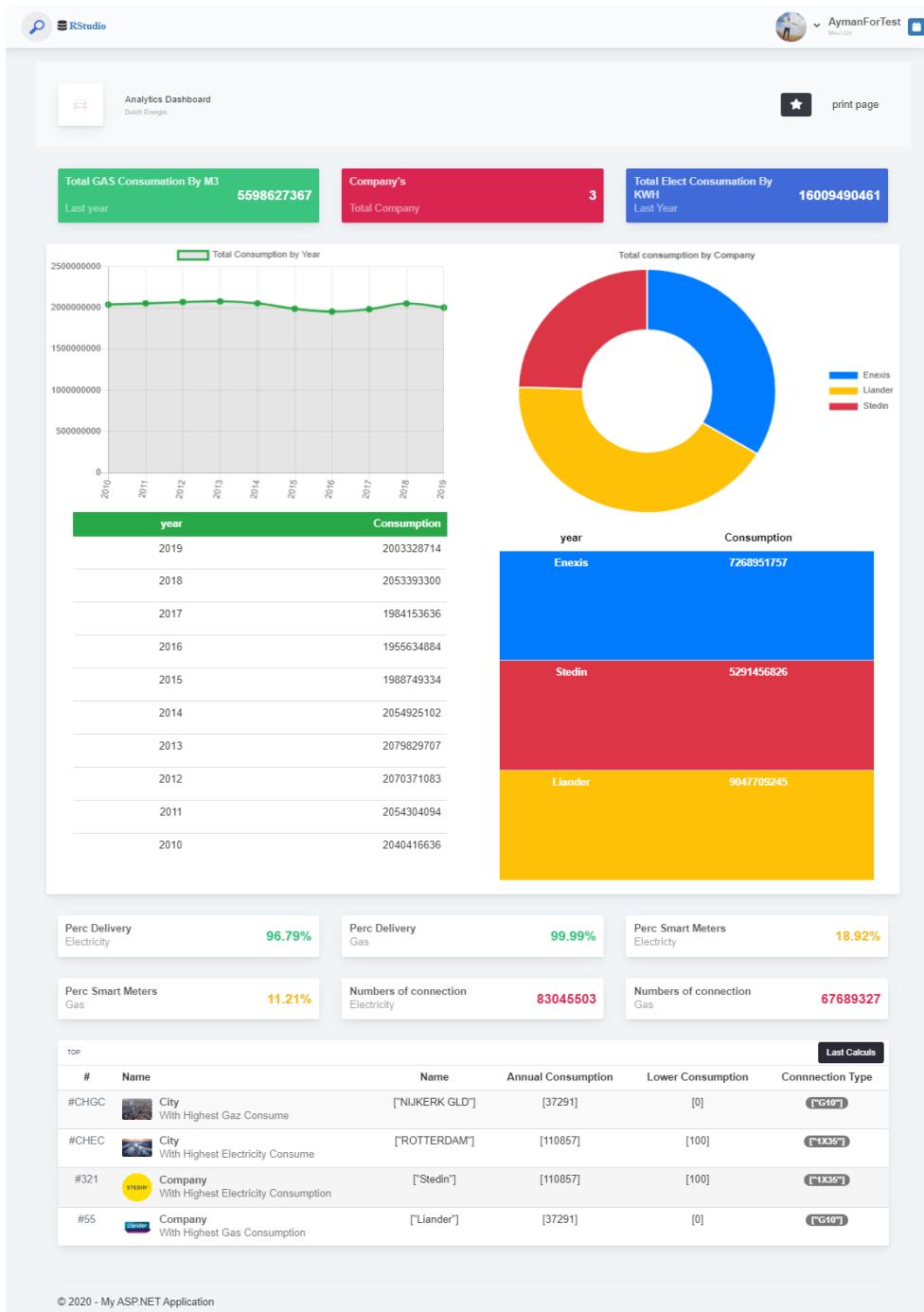
ASP.NET est un framework permettant de générer à la demande des pages web, lancé par Microsoft en juillet 2002, et utilisé pour mettre en œuvre des applications web3.

3.3 Restitution de donnée

La dernière phase de notre projet, et qui constitue en l'occurrence la dernière étape de tout projet décisionnel, consiste en l'élaboration des Tableaux de bord qui offriront aux utilisateurs finaux la possibilité d'accéder de manière autonome et interactive aux informations.

Les tableaux de bord que nous allons présenter au client contiennent des tableaux, matrices, graphes.

3.3.1 Tableaux de Bord 'Dashbord'



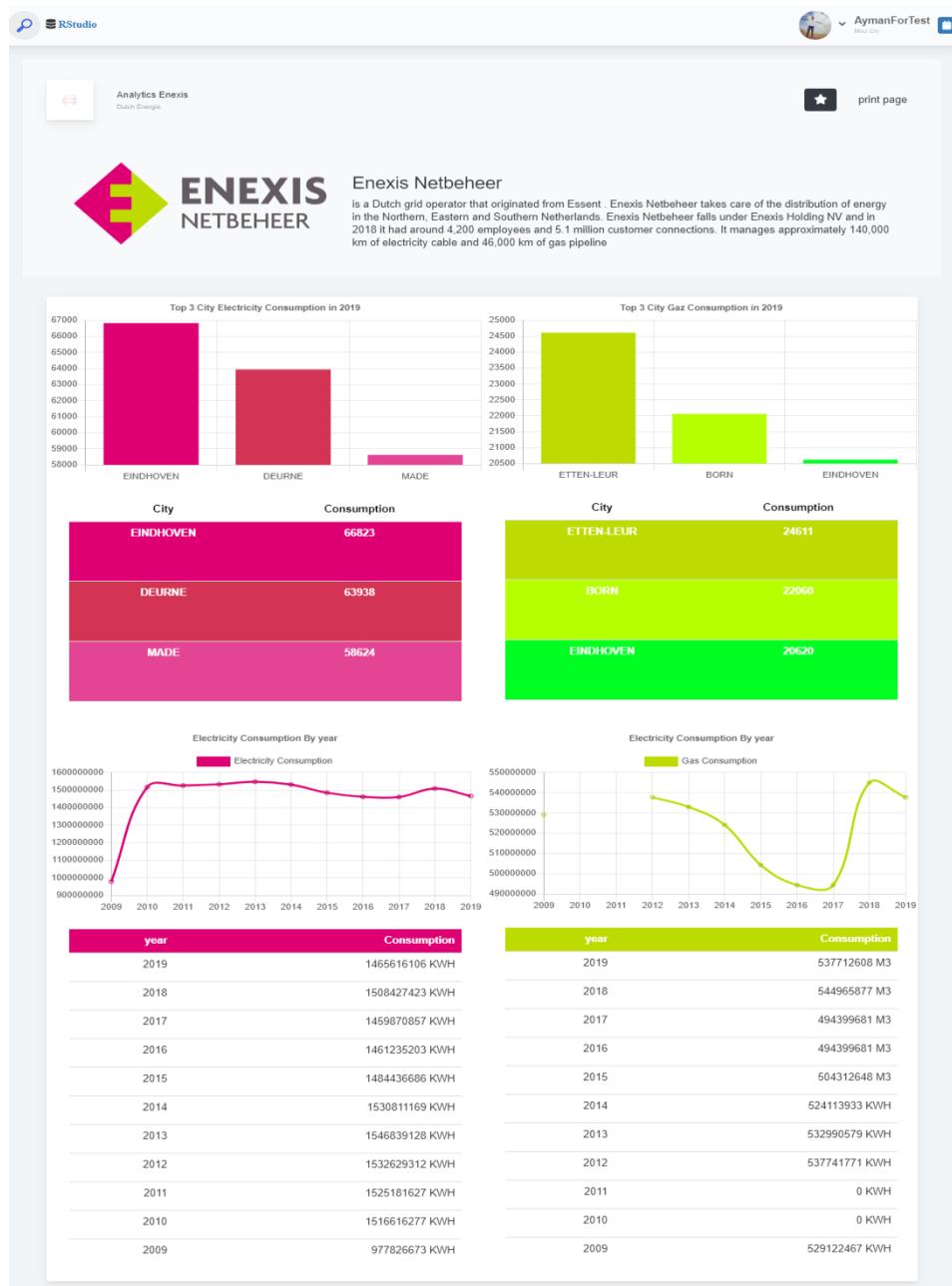
Figures 33: Tableaux de bord 'Dashbord'

Ce tableau de bord contient les nombres total des sociétés, consommation électrique depuis 2009 et consommation gaz depuis 2019 expose deux graphes la consommation total de l'année 2019 pour chaque société et aussi la tendance de la consommation depuis 2009 les deux graphes sont présentés aussi sous des tableaux pour que les données soient lisibles leur d'impression

Ainsi que j'ai intégré 6 labels qui nous montre le pourcentage d'énergie distribuée le nombre de connexion et le pourcentage des compteurs intelligent pour chaque catégorie

Et en fin j'ai ajouté un tableau qui affiche la ville et la sociétés les plus consommateur pour chaque catégories

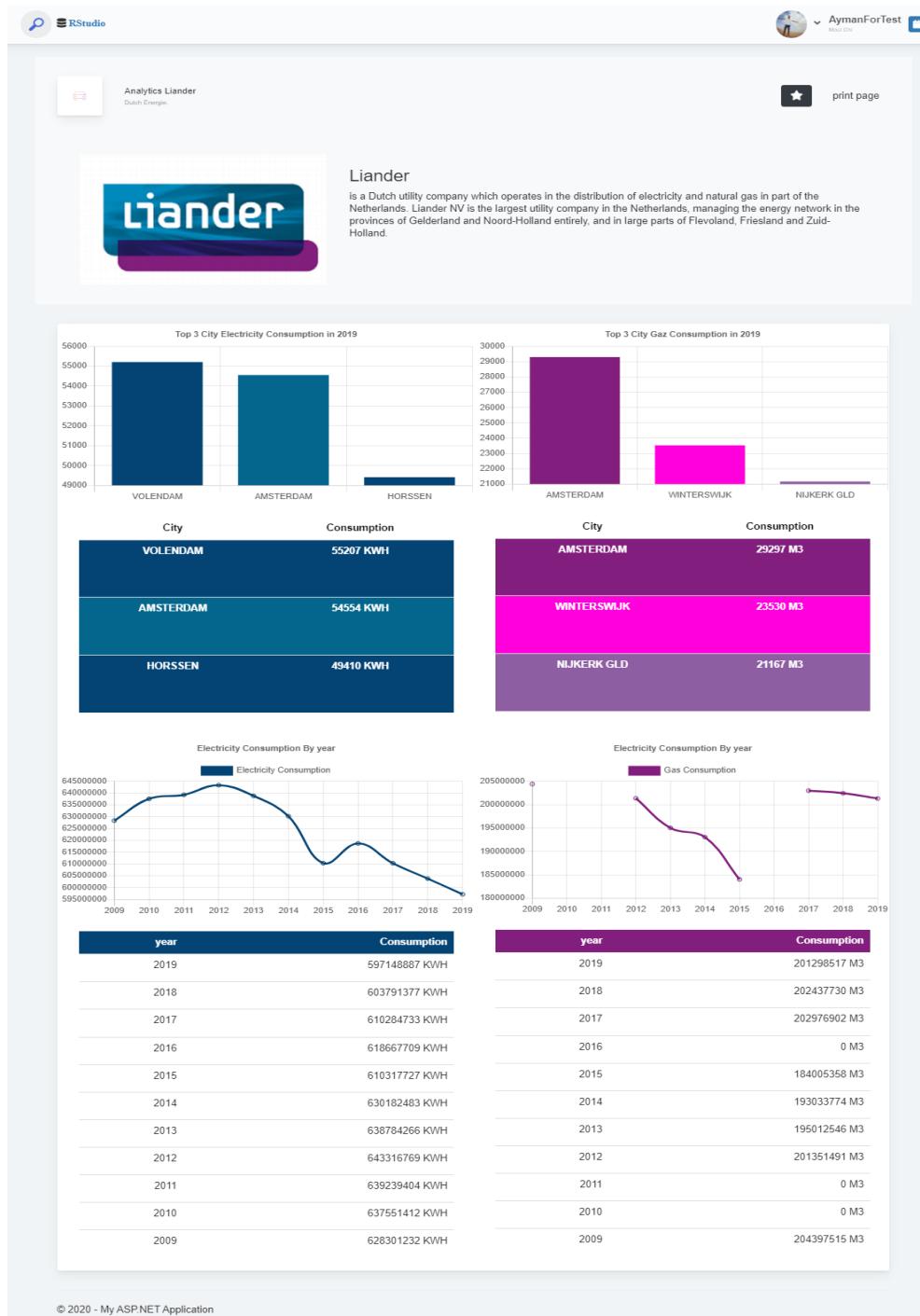
3.3.2 Tableaux de bord 'ENEXIS'



Figures 34: Tableau de bord 'ENEXIS'

Ce tableau de bord contient deux diagramme à bandes et deux graphique en ligne, exposant les trois ville les plus consommateur pour chaque catégorie avec le total de la consommation et la tendance de consommation pour chaque catégorie tous les graphes son aussi affiche sous des tableau pour que dans l'impression les données soient lisibles

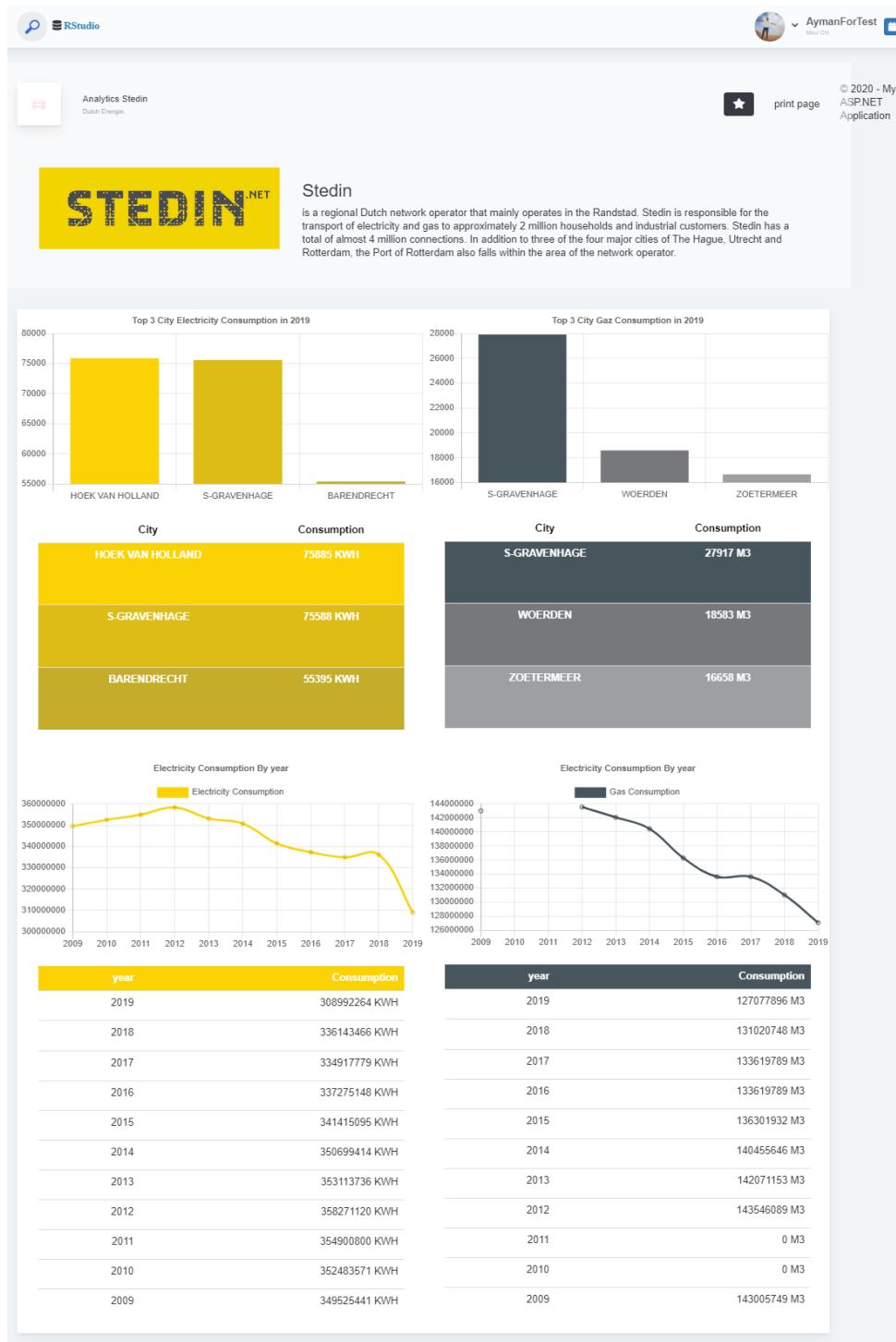
3.3.3 Tableaux de bord 'LIANDER'



Figures 35: Tableau de bord 'LIANDER'

Ce tableau de bord contient deux diagramme à bandes et deux graphique en ligne, exposant les trois villes les plus consommateur pour chaque catégorie avec le total de la consommation et la tendance de consommation pour chaque catégorie tous les graphes son aussi affiche sous des tableaux pour que dans l'impression les données soient lisibles

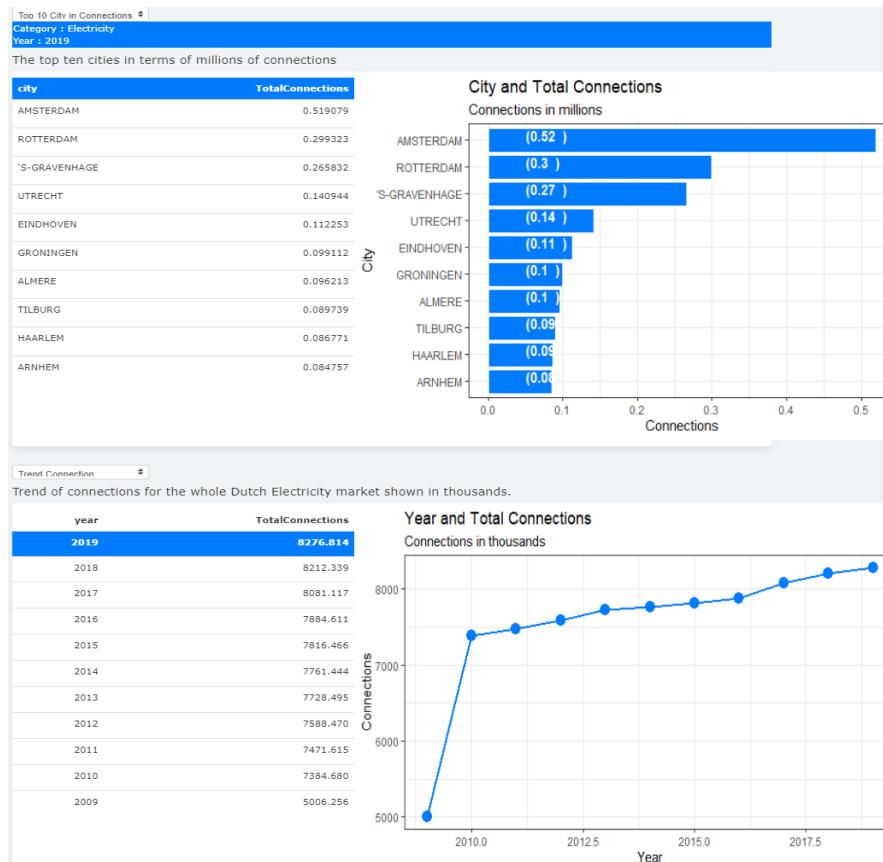
3.3.4 Tableaux de bord 'LIANDER'

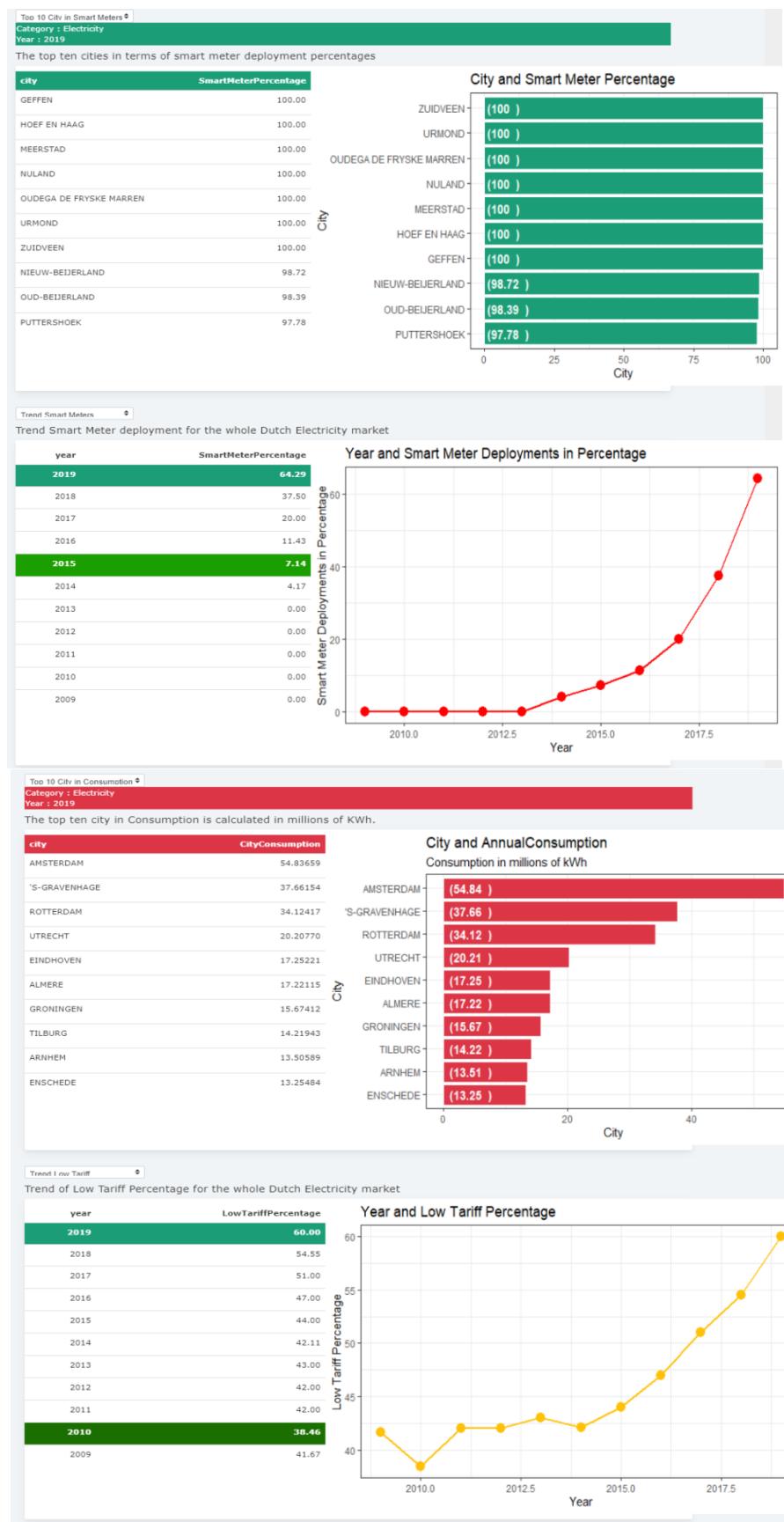


Figures 36: Tableau de bord 'LIANDER'

Ce tableau de bord contient deux diagramme à bandes et deux graphique en ligne, exposant les trois ville les plus consommateur pour chaque catégorie avec le total de la consommation et la tendance de consommation pour chaque catégorie tous les graphes son aussi affiche sous des tableau pour que dans l'impression les données soient lisibles

3.3.5 Tableaux de bord 'RStudio'





Figures 37: Tableau de bord ‘RSTUDIO’

Ce tableaux de bord contient trois graphes en ligne et trois tableaux, exposant la tendance de (Tarif Bad – Annual Consommation – Compteurs Intelligents Connection)

5. Conclusion

Ce chapitre a concerné l'aboutissement au résultat attendu de notre projet à savoir la création d'un système décisionnel. Il a détaillé d'abord le processus ETL que nous avons élaboré avec SSIS. Ce processus est une composante clé de notre plateforme vu qu'il assure l'alimentation du datawarehouse conçu et qu'il prévoit son évolutivité. Ce datawarehouse a été ensuite exploité pour la création et le déploiement de nos cubes OLAP et nos Tableau de bord. Les cubes OLAP présentent un outil de pilotage et d'aide à la décision dans la mesure où ils permettent selon des techniques appropriées de navigation un filtre sur les données à générer, donc une information précise, et offrent des tableaux croisés dynamiques pour d'éventuels besoins avancés d'analyse. Les rapports statiques comportent des graphes et des tableaux croisés qui permettent à l'utilisateur une vue globale mais aussi synthétique des données. Dans ce cadre, Nous avons donné, dans ce chapitre, quelques exemples de cubes OLAP et de rapports générés.

VI. Analyse financière

1. Evaluation financière

L'évaluation financière peut être définie comme une démarche, qui s'appuie sur l'examen critique de l'information comptable et financière fourni par une entreprise à destination des tiers, ayant pour but d'apprécier de plus objectivement possible sa performance financière et économique (rentabilité, solvabilité, risque). Dont les outils sont la valeur actuelle, l'indice de profitabilité, le taux de rentabilité interne et le délai de récupération de capital investi.

1.1 Valeur actuelle nette ou VAN

La valeur actuelle nette est une valeur de revenu futur actualisée à un coût approprié. La VAN est la différence entre la marge brute d'autofinancement actualisée cumulée et l'investissement initial (I)

$$VAN = \sum_{n=1}^N C_n (1+i)^{-n} - I$$

Avec : VAN : la valeur

actuelle nette C

: cash – flow

n : nombre d'années exprimant la durée de vie du projet

i : taux d'actualisation (taux

de référence) I : capital investi

(58 887 900 MAD)

	A	B	C	D	E	F	G
1		Investissement	Invest 2	Recettes	Charges	Flux Non Actualisé	Actualisaton a TRM = 4%
2 T=0		300,000.00 MAD		200,000.00 MAD	180,000.00 MAD		
3 T=1				300,000.00 MAD	240,000.00 MAD		
4 T=2				336,000.00 MAD	258,000.00 MAD		
5 T=3			500,000.00 MAD	376,320.00 MAD	277,350.00 MAD		
6 T=4				421,478.40 MAD	298,151.25 MAD		
7 T=5				472,055.81 MAD	320,512.59 MAD		

Par la formule de calcul de la VAN :

Annee	T0	T1	T2	T3	T4	T5
Cash-Flow	100 000	150000	168000	188160	210739	236027.91
I (Capital Investi)	300 000	-	-	500 000	-	-
Taux actualisation	0.714	0.510	0.364	0.714	0.510	0.364
Cash Flow Actual	71,400	76,500	61,152	134,346.24	107,476.89	85,914.15
CF Cumule	71,400	147,900	209,052	343,398.24	450,875.13	536,789.28

Détermination du cash - flow

$$\text{VAN (3ans)} = 343\,398.24 - 300\,000$$

$$\text{VAN (3ans)} = 43\,398.24 \text{ MAD}$$

$$\text{VAN (5ans)} = 536\,789.28 - 500\,000$$

$$\text{VAN (5ans)} = 36\,789.28 \text{ MAD}$$

Par interprétation :

La Van permet de mesurer la création de richesses attendues d'un projet d'investissement. La VAN est positif pour (3 ans et pour 5ans). Donc le projet est rentable, car les capitaux investis peuvent être récupérés et dégage un surcroit de richesse.

1.2 Indice de rentabilité IR ou indice de profitabilité IP

L'indice de profitabilité est le rapport entre la somme des marges brutes actualisées et la somme de capitaux investis. L'IP sert à apprécier la faculté de l'entité à donner des profits. Il présente le profit généré par 1 MAD investi.

L'indice de profitabilité peut être calculé de deux méthodes :

Indice de profitabilité = Somme des flux actualisés / investissement initial

Indice de profitabilité = (VAN / I (capital investi)) + 1

VAN (3ans) = 43 398.24 MAD

VAN (5ans) = 36 789.28 MAD

capital investi = 300 000 MAD

L'IP calculée par la formule suivante :

$$IP = \frac{\sum C_n (1+i)^{-n}}{I}$$

$$IP (5 \text{ ans}) = (36 789.28 / 300 000) + 1$$

$$IP = 1, 07$$

$$IP (3 \text{ ans}) = (43 398.24 / 300 000) + 1$$

$$IP = 1, 14$$

L'indice de profitabilité est de 1,14 MAD. Ceci explique que l'investissement par la réalisation du projet permettra de recevoir 1,14 MAD sur 1 MAD investi. L'opération est donc avantageuse car l'indice de profitabilité est supérieur à 1 et la VAN positive affirme que le projet acceptable.

1.3 Taux de rentabilité interne ou TRI

Le taux de rentabilité interne, est le taux qui annule la valeur actuelle nette de cette entreprise, il y a donc une équivalence entre le capital investi et l'ensemble des cash-flows réalisés dans la durée de vie du projet

Détermination du TRI

Une formule algébrique possible pour le TRI est:

$$\text{TRI} = R1 + [V1(R2 - R1)/(V1 - V2)]$$

Où:

- **R1 ,R2** = taux d'actualisation choisis au hasard
- **V1** = valeur actuelle nette plus élevée
- **V2**= valeur actuelle nette plus basse

Figures 38: Formule de calcul de TRI

TRI = 5,28 %

Le taux interne de rendement est égal à 5,28 % ce qui est largement au taux d'intérêt qui est fixé à 4%, donc le projet est rentable et bancable avec une marge de sécurisée de 1,28%

Délai de récupération de première investissement

DR = 2.67 = (2 ans 8 mois)

Délai de récupération de deuxième investissement

DR = 4.57= (4 ans 6 mois)

Conclusion générale

Notre projet consistait à mettre en œuvre un système décisionnel pour la fonction energie. Il s'agit précisément de générer des tableaux de bord qui offrent une meilleure visibilité et qui mettent à disposition l'information essentielle pour la prise de décision relativement aux thème energie.

Vu le caractère décisionnel du projet, nous avons suivi les différentes étapes du cycle de vie dimensionnel. Après avoir étudié l'existant et recueilli les besoins des utilisateurs finaux, nous avons défini les indicateurs et les axes d'analyse et abouti à l'architecture technique de la solution. Grâce aux indicateurs et axes d'analyse identifiés, nous avons pu dresser une conception du modèle dimensionnel du datawarehouse et de la présentation des tableaux de bord. Nous nous sommes ensuite consacrés à la réalisation de la solution en alimentant le datawarehouse grâce à l'outil SSIS et en générant les états de restitution ciblés sous forme d'un application web des tableaux de bord sur la base de cubes OLAP avec SSAS ou de Tableau de bord à l'aide de ASP MVC ET CHARTJS ET RSTUDIO.

Au terme de ces étapes, nous pouvons affirmer que notre plateforme a satisfait les clauses du cahier de charges de notre projet et permet des possibilités intéressantes d'analyse et d'accès à l'information offertes aux utilisateurs finaux.

Par ailleurs, ce projet m'a été bénéfique aussi bien au niveau informatique qu'au niveau professionnel. Il m'a permis de raffiner mes capacités d'abstraction et de conception en modélisation décisionnelle et offert l'occasion de découvrir et manipuler plusieurs outils Open Source de la BI et enfin de m'ouvrir plus sur l'énergie.

En perspective, on peut envisager d'intégrer d'autres thèmes de la fonction Energie dans notre système décisionnel.

Bibliographie/Webographie

Bibliographie

[Bouquin, 2003] : Bouquin Henry ; « Le contrôle de gestion » ; P.U.F ; 2003.

[Dresner, 2001] : H. Dresner ; « BI : Making the Data Make Sens » ; Gartner Group 2001.

[Franco, 1997] : Jean-Michel Franco; « Le Data Warehouse, le Data Mining » ; Eyrolles 1997.

[Goglin, 1998] : J.F. Goglin; « La Construction du Datawarehouse : du Datamart au Dataweb »; Hermes 1998.

[Inmon, 2002]: W. H. Inmon ; « Building the Data Warehouse Third Edition» ; Wiley Computer Publishing 2002.

[Kimball, 2004] : R. Kimball et J. Caserta ; « The Data warehouse ETL Toolkit» ;Wiley Publisshing, INC 2004

[Kimball, 2002] : R. Kimball et M. Ross ; « Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition » ; Vuibert 2002.

[Kimball,1996] : R. Kimball ; « Entrepôts de données : Guide pratique du concepteur de Data Warehouse » ;Wiley Computer Publishing 1996.

[Le Moigne, 1977] : Le Moigne J.L., « La théorie du système général, théorie de la modélisation », P.U.F., 1977.

[Nakache, 1998] : Didier Nakache; « Data Warehouse et Data Mining »; Conservatoire National des Arts et Métiers de Lille; Version 1.1; 15 juin 1998.

[Bouzghoub, 2008] : Abdenour Bouzghoub ; « Modélisation des Entrepôts de données XML : Application au domaine de la sécurité sociale » ; Thèse de Magistère Option : SISCSD ; Institut National de Formation en Informatique (I.N.I) 2008.

[Chouder, 2007] : Lamri Chouder ; « Entrepôt Distribué de Données » ; Thèse de

Magistère Option : SI; Institut National de Formation en

Informatique (I.N.I) 2007.

[Chuck, 1998] : Chuck Ballard, Dirk Herreman, Don Schau, Rhonda Bell, Eunsaeng

Kim, Ann Valencic; Data Modeling Techniques for Data

Warehousing; International Technical Support Organization;

<http://www.redbooks.ibm.com>; février 1998.

Webographie

<https://decizia.com/cycle-de-vie-decisionnel-definition/>

<https://docs.microsoft.com/en-us/sql/ssdt/download-sql-server-data-tools-ssdt?view=sql-server-ver15>

<https://www.kaggle.com/lucabasa/dutch-energy>

<https://www.chartjs.org/samples/latest/>

<https://community.rstudio.com/>

https://stackoverflow.com/*

<https://www.researchgate.net/>

<https://www.ieee.org/>

Annexe

A

Alimentation	6, 13, 23, 32, 47, 48, 50, 52, 66, 71 73
Architecture technique	20, 71
Axe	25, 26, 27, 28, 29, 30, 31, 34, 35, 36, 37, 38, 43, 44, 45, 54, 59, 61, 74, 75

B

Besoins	5, 18, 19, 20, 22, 23, 24, 26, 27, 34, 35, 36, 37, 38, 45, 57, 63, 70, 71, 74
Business Intelligence	6, 7, 13, 33, 74, 76

C

Cadrage	13, 24, 27, 32
Cahier de charges	22, 26, 74
Conception	14, 16, 17, 19, 20, 22, 25, 33, 34, 35, 36, 43, 45, 71
Conditions de travail	13, 19, 23, 27, 30, 34, 41, 45, 67, 81, 83
Cycle de vie	6, 14

D

Datamart	5, 6, 36, 38, 39, 40, 41, 45, 74, 75
Datawarehouse	5, 6, 22, 23, 36, 45, 46, 47, 54, 70, 71, 72, 74
Dimension	7, 22, 31, 33, 38, 47, 48, 49, 52, 56, 57, 71, 74, 75

E

Energie	27, 28, 30, 37, 38, 43, 44, 45, 57, 60, 62, 63, 64, 82, 92
ETL	7, 32, 33, 46, 47, 70, 71, 75

F

Fait	38, 41, 50, 52, 53, 56, 57, 74, 75
------	------------------------------------

I

Indicateur	19, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 43, 44, 45, 74, 75
------------	--

M

Maintenance	19, 20, 22
Modélisation	20, 22, 36, 38, 71, 75

O

OLAP 5, 6, 7, 14, 22, 33, 46, 54, 58, 70, 71, 75

Outil 5, 13, 18, 20, 23, 25, 32, 33, 34, 47, 54, 63, 70, 71, 73, 75

P

Plan Assurance Qualité 7, 13, 15, 22, 23

Planning 13, 15, 18, 20, 22, 23

Plateforme décisionnelle 5, 13, 20, 33, 65, 70

R

Règles de gestion 28, 29, 30, 31

Rémunération 5, 13, 19, 23, 27, 29, 34, 37, 41, 44, 45, 63, 64, 71

Reporting 5, 6, 7, 14, 24, 25, 26, 33, 34, 71, 72

S

SGBD 7, 20, 25, 33, 83

Sources de données 5, 24, 25, 26, 33, 36, 47

SQL 7, 33, 53, 72

T

Tableaux de bord

5, 43, 44, 45