



**HAL**  
open science

# Comment une solution décisionnelle peut-elle permettre aux gestionnaires de disposer de données provenant de sources métiers hétérogènes ?

Fabien Pelat

## ► To cite this version:

Fabien Pelat. Comment une solution décisionnelle peut-elle permettre aux gestionnaires de disposer de données provenant de sources métiers hétérogènes ?. Recherche opérationnelle [math.OA]. 2016. dumas-01876443

**HAL Id: dumas-01876443**

**<https://dumas.ccsd.cnrs.fr/dumas-01876443>**

Submitted on 18 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**CONSERVATOIRE NATIONAL DES ARTS ET METIERS**

**CENTRE REGIONAL ASSOCIE DE MIDI PYRENEES**

---

**MEMOIRE**

**présenté en vue d'obtenir**

**le DIPLOME D'INGENIEUR CNAM**

**SPECIALITE : INFORMATIQUE**

**OPTION : Système d'information**

**par**

**Fabien PELAT**

---

**Comment une solution décisionnelle peut-elle permettre aux gestionnaires de disposer de données provenant de sources métiers hétérogènes ?**

**Soutenu le 2 Juin 2016**

---

**JURY**

**PRESIDENT : YANN POLLET**

**MEMBRES : THIERRY MILLAN  
XAVIER CREGUT**



## **Remerciements**

Je tiens à remercier l'ensemble des personnes : supérieurs, collègues et stagiaires, avec qui j'ai collaboré tout au long de ce projet ; ainsi que la direction d'EDENIS pour m'avoir confié ce projet.

Je souhaite tout particulièrement remercier Monsieur Jérôme RIO, responsables contrôle de gestion et système d'information, pour ses remarques pertinentes, ses précieux conseils et son exigence.

Je tiens également à remercier l'institution de l'IPST CNAM pour la qualité de ses enseignements et l'opportunité qu'elle m'a offerte de poursuivre mon parcours de formation en parallèle de ma carrière.

Je n'aurais probablement pas réalisé ce mémoire sans le soutien de mes proches, parents et amis que j'associe à ce travail, et tout particulièrement mon épouse Virginie qui a été un soutien permanent malgré mes absences.

## Liste des abréviations

- BI : « Business intelligence ». Ensemble de données consolidées qui permet la prise de décision.
- CSV : Format de fichier contenant des valeurs séparées par des virgules.
- DSI : Direction des systèmes d'information.
- EHPAD : Etablissement d'Hébergement pour Personnes Agées Dépendantes
- ERP : « Enterprise Resource Planning ». Progiciel de gestion intégré (PGI).
- ETL : « Extract Transform Load. »
- ETP : Equivalent temps plein
- GIR : Groupe Iso ressource
- IPP : identifiant personnel patient
- MPLS : « MultiProtocol Label Switching «
- ODS : « Operational Data Store ». Zone de préparation des données.
- OLAP : « Online Analytical Processing ». Traitement analytique en ligne.
- OLTP : « online transaction processing ». Traitement transactionnel en ligne.
- SAAS : Software as a service
- SCII : Société de Conseil en Ingénierie Informatique
- SDSI : Schéma directeur des systèmes d'information
- SGBD : Système de Gestion de bases de données.
- SID : Système d'information décisionnel
- SIG : Système d'information de gestion
- SIO : Système d'information opérationnel

# Glossaire

- **Datamart :** Sous ensemble thématique d'un Datawarehouse.
- **BI :** La business intelligence ou informatique décisionnelle désigne les outils et méthode destiné à l'exploitation des données de l'entreprise dans un objectif de prise de décision.
- **BIG DATA :** Désigne l'ensemble des données numériques produites par l'utilisation des nouvelles technologies. Cela recoupe les données d'entreprise aussi bien que des données issues de capteurs, des contenus publiés sur le web (, des transactions de commerce électronique, des échanges sur les réseaux sociaux, des données transmises par les objets connectés, des données géo localisées, etc.
- **Data warehouse :** Un entrepôt de données est une base de données dans laquelle les informations stratégiques de différentes sources de données sont synthétisées et réorganiser pour en faciliter l'analyse selon différentes axes prédéterminés.
- **ETL :** « Extract Transform Load ». Processus ayant pour but de récupérer les données des bases de production pour les injecter dans le Data warehouse après avoir effectué des transformations.
- **Métadonnées :** Ensemble structuré de données créées pour fournir des informations sur des ressources électroniques.
- **ODS :** Base de données conçus pour centraliser les données issues de sources hétérogènes afin de faciliter les opérations d'analyse et de reporting.
- **SAAS :** Les Logiciels en tant que Service désignent des logiciels qui sont hébergés sur le serveur d'un prestataire, accessibles à distance, et dont la facturation s'effectue sous forme d'abonnement, ou proportionnellement à l'utilisation de certaines ressources.
- **SGBD** Un SGBD est un logiciel qui prend en charge la structuration, le stockage, la mise à jour et la maintenance d'une base de données.

# Table des matières

<b>REMERCIEMENTS</b> .....	<b>3</b>
<b>LISTE DES ABREVIATIONS</b> .....	<b>4</b>
<b>GLOSSAIRE</b> .....	<b>5</b>
<b>TABLE DES MATIERES</b> .....	<b>6</b>
<b>I INTRODUCTION</b> .....	<b>8</b>
I.1 PRESENTATION GENERALE D'EDENIS.....	8
I.1.1 <i>L'association</i> .....	8
I.1.2 <i>Les résidences</i> .....	10
I.2 PRESENTATION DU PROJET .....	11
I.3 PRESENTATION DU SERVICE SYSTEME D'INFORMATION D'EDENIS.....	12
I.3.1 <i>Organisation humaine</i> .....	12
I.3.2 <i>Cartographie Système d'information</i> .....	13
<b>II ETAT DE L'ART</b> .....	<b>15</b>
II.1 LES ORIGINES .....	15
II.2 LES BESOINS EN PILOTAGE .....	15
II.3 LES THEORICIENS .....	18
II.4 CONCEPTS DECISIONNEL .....	19
II.4.1 <i>Modélisation des informations d'aide à la décision</i> .....	19
II.4.2 <i>Concept de tables de fait</i> .....	19
II.4.3 <i>Concept de dimensions</i> .....	21
II.4.4 <i>La modélisation en étoile</i> .....	23
II.4.5 <i>La modélisation en flocon</i> .....	23
II.4.6 <i>La modélisation en constellation</i> .....	23
II.5 ARCHITECTURE D'UN SYSTEME DECISIONNEL .....	24
II.5.1 <i>Composant de collecte</i> .....	24
II.5.2 <i>Composant de stockage</i> .....	25
II.5.3 <i>Composant de diffusion</i> .....	25
II.5.4 <i>Composant d'exploitation</i> .....	25
II.6 MISE EN ŒUVRE D'UN PROJET DECISIONNEL .....	26
II.6.1 <i>Les préalables projet</i> .....	26
II.6.2 <i>Le sponsoring</i> .....	26
II.6.3 <i>L'intérêt métier</i> .....	26
II.6.4 <i>La quantité et la qualité des données</i> .....	27
II.6.5 <i>La sélection des domaines à traiter</i> .....	28
II.6.6 <i>Le cycle de vie d'un projet décisionnel</i> .....	29
II.7 LE MARCHE DU DECISIONNEL .....	32
II.7.1 <i>Historique</i> .....	32
II.7.2 <i>Les tendances</i> .....	33
<b>III PROJET DECISIONNEL EDENIS</b> .....	<b>35</b>
III.1 ORIGINES DU PROJET.....	35
III.1.1 <i>Les besoins à l'origine du projet</i> .....	35
III.1.2 <i>Historique des projets précédents</i> .....	36
III.1.3 <i>Pré-étude du nouveau projet</i> .....	37
III.1.4 <i>Proposition projet</i> .....	40
III.2 PROOF OF CONCEPT .....	41
III.2.1 <i>Analyse de l'existant</i> .....	41
III.2.2 <i>Source de données</i> .....	42
III.2.3 <i>Règles de gestion</i> .....	43

III.2.4	Paramètres de mises à jour et de diffusion des rapports .....	45
III.2.5	Installation et développement.....	45
III.2.6	Tests.....	45
III.3	GESTION DE PROJET .....	47
III.4	PHASE 0 : INITIALISATION PROJET.....	49
III.4.1	Mise en place infrastructure et environnement de développement .....	49
III.5	PHASE 1 : SELECTION D'UN DOMAINE .....	50
III.6	PHASE 2 : DEFINITION DES BESOINS.....	50
III.7	PHASE 3 A : MODELISATION DES DONNEES. ....	51
III.7.1	Analyse des données existantes. ....	51
III.7.2	Modélisation ODS .....	53
III.7.3	Conception des ETL ODS.....	53
III.7.4	Modélisation DATA WAREHOUSE.....	54
III.7.5	Conception ETL DWH.....	55
III.8	PHASE 3 B : CONCEPTION DES COMPOSANTS DE LA SOLUTION DECISIONNELLE AFIN DE REpondre AUX PROBLEMATIQUES METIERS. ....	56
III.8.1	Problématique de fragmentation du système d'information.....	56
III.8.2	Problématique d'indépendance au système d'information et historisation des données .....	58
III.8.3	Problématique de consolidation des données.....	59
III.8.4	Comment résoudre la problématique de mise à disposition des données aux utilisateurs ? ...	60
III.8.5	Comment répondre au besoin d'autonomie pour la création de rapports ?.....	62
III.8.6	Comment répondre à la problématique de diffusion des rapports par mail ? .....	65
III.8.7	Comment répondre à la problématique d'intégration de données non structurées ? .....	67
III.8.8	Résultats.....	67
III.8.9	Comment déclencher manuellement un package ETL ?.....	68
III.9	PHASE 3 B : CONCEPTION DES COMPOSANTS DE LA SOLUTION DECISIONNELLE AFIN DE REpondre AUX PROBLEMATIQUES D'EXPLOITATION. ....	69
III.9.1	Gestion des erreurs ETL ? .....	69
III.9.2	Comment maîtriser la volumétrie et les performances ?.....	70
III.9.3	Comment gérer les autorisations ? .....	73
III.10	EXEMPLE DE MISE EN ŒUVRE D'INTEGRATION DANS L'ODS .....	75
III.10.1	Sélection d'un domaine. ....	75
III.10.2	Définition des besoins .....	75
III.10.3	Modélisation. ....	75
III.10.4	Conception des composants.....	76
III.11	EXEMPLE DE MISE EN ŒUVRE D'INTEGRATION DANS LE DATA WAREHOUSE.....	78
III.11.1	Sélection d'un domaine. ....	78
III.11.2	Définition des besoins .....	78
III.11.3	Modélisation. ....	78
III.11.4	Conception des composants.....	80
III.12	EVOLUTIONS DU PROJET : .....	83
III.12.1	Prise en compte des applications SAAS.....	83
III.12.2	Intégration aux outils bureautique SAAS.....	84
III.12.3	Evolution du système d'information.....	86
III.12.4	Autres perspectives .....	86
<b>CONCLUSION.....</b>		<b>87</b>
	BILAN DE LA PROPOSITION PROJET : .....	87
	BILAN DES OBJECTIFS DU PROJET.....	88
	BILAN HUMAIN .....	88
<b>BIBLIOGRAPHIE .....</b>		<b>91</b>
<b>TABLE DES ANNEXES .....</b>		<b>92</b>
<b>LISTE DES FIGURES.....</b>		<b>102</b>
<b>LISTE DES TABLEAUX.....</b>		<b>103</b>

# **I Introduction**

## **I.1 Présentation Générale d'EDENIS**

### **I.1.1 L'association**

Edenis est un groupement associatif de 18 Etablissement (Annexe 1) en région Midi-Pyrénées. Le groupe accueille plus de 1600 résidents, dont 179 en unité de vie protégée.

L'association Promo Accueil est créée en 1985. En 2013 elle devient Edenis et témoigne depuis 30 ans de son implication dans l'accompagnement du grand âge.

Edenis et ses 1500 collaborateurs ont pour volonté d'être innovants dans la prise en charge du résident. Au cœur d'une démarche d'amélioration de la continuité de soins, le groupe Edenis est toujours plus engagé dans la prévention de la dépendance et le maintien des capacités de ses résidents.

Edenis s'appuie sur 3 axes stratégiques pour devenir un acteur de référence :

- L'établissement, lieu d'excellence opérationnelle
- Le résident au centre des préoccupations
- Le collaborateur comme principale richesse

Le partenariat avec le laboratoire de Géronto-Technologie de Toulouse La Grave fait d'Edenis un groupe avant-gardiste, proposant de nouvelles solutions innovantes pour améliorer la vie et le confort du résident.

Edenis est engagé dans une démarche qualité qui se concrétise notamment par l'obtention de la certification Qualicert SGS en 2014 et renouvelée pour 2015.

L'association EDENIS est composée d'environ 1550 professionnels répartis sur les différents établissements du groupe et répartis selon les professions suivantes :

- 350 Médecins Généralistes ;
- 20 Médecins Coordinateurs
- 20 Psychologues salariés
- 100 Infirmières Diplômées d'Etat & IDE cadre
- 700 Aides-Soignants, Assistants Médico-Psychologiques, Agents de Service Polyvalent, Faisant fonction & agents d'accompagnement
- 250 Intervenants extérieurs (ergothérapeute, kinésithérapeute, pédicure, orthophoniste, dentiste, gériatre, diététicien, ...)
- 20 animateurs projet de vie
- 30 Directeurs d'établissement et directeurs adjoints
- 50 Agents d'accueil, assistants administratifs

La gouvernance de l'association est présentée dans la figure 1.

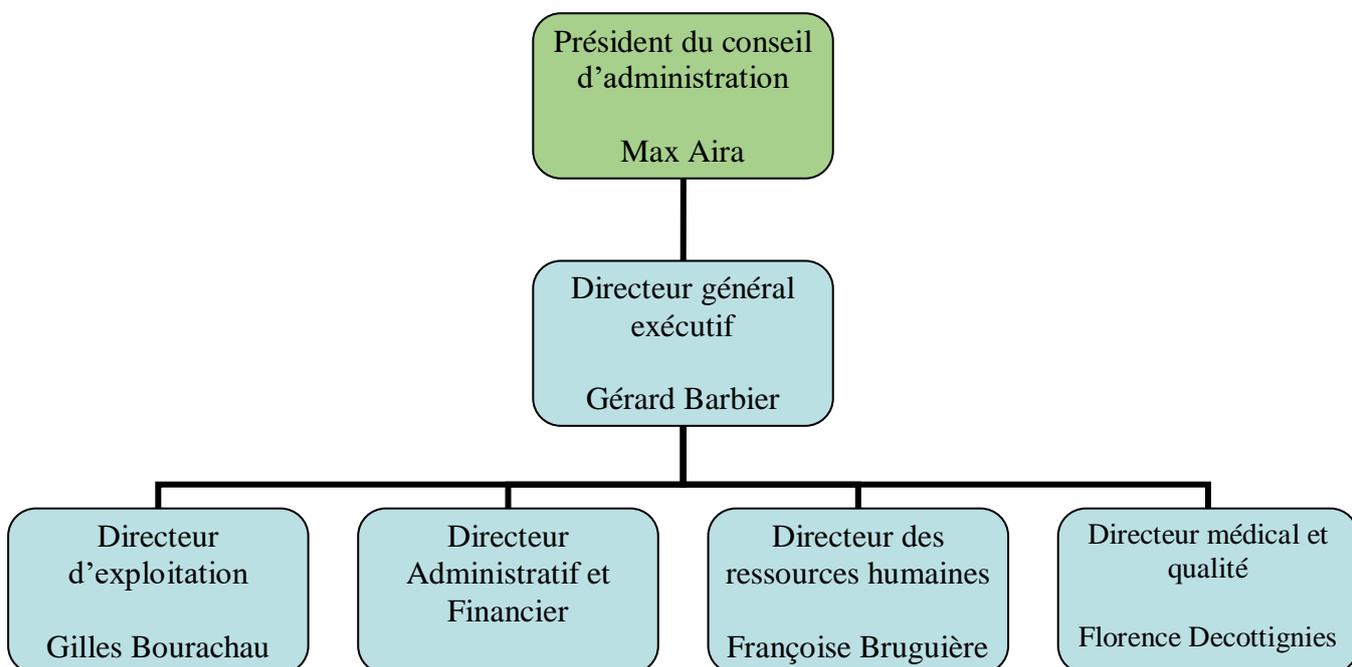


Figure 1 Organigramme Direction générale EDENIS

## I.1.2 Les résidences

Les résidences EDENIS sont toutes des Etablissements d’Hébergement pour Personnes Agées Dépendantes (EHPAD) destinées à accueillir des personnes seules ou en couple âgées de plus de 60 ans (sauf dérogation), plus ou moins dépendantes sur les plans physique et psychologique. La vie en couple est maintenue grâce aux chambres doubles prévues à cet effet.

La dépendance se définit comme la perte d’autonomie et l’incapacité pour une personne de réaliser, sans aide extérieure, les gestes essentiels de la vie quotidienne : préparer à manger, se laver, s’habiller, faire des courses, etc. Cette perte d’autonomie peut être psychique, physique ou les deux à la fois.

Elle est évaluée au travers d’une grille élaborée pour déterminer le degré de perte d’autonomie physique et/ou psychique d’une personne âgée, et ainsi définir ses besoins. Selon le niveau de dépendance, les personnes sont classées en six groupes iso-ressources (GIR). Sont qualifiées de dépendantes les personnes en GIR de 1 à 4, les personnes en GIR 5 et 6 étant très peu ou pas dépendantes.

Chaque résident bénéficie dans les semaines suivant son arrivée, d’une évaluation gériatrique standardisée (statut nutritionnel, autonomie, risque de chute, risque de dépression, statut cognitif). Ces données, associées au recueil de ses habitudes de vie, de ses goûts et de ses loisirs à domicile permettent de lui proposer, sans jamais imposer, un accompagnement personnalisé respectueux de ses attentes et adaptés à ses besoins et capacités. Ce projet personnalisé, validé par le résident et sa famille, est révisé chaque année et à chaque fois que son état de santé le nécessite.

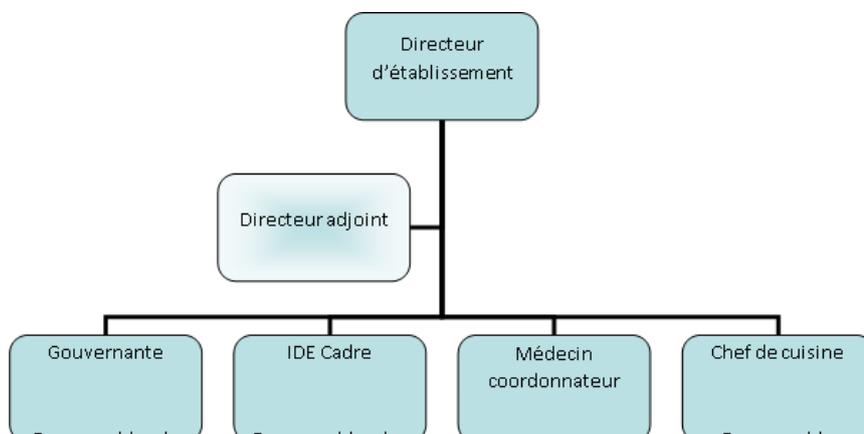


Figure 2 Organigramme EHPAD

## I.2 Présentation du projet

La direction financière au travers du service de contrôle de gestion est en charge de la réalisation de l'ensemble des rapports et tableaux de bords émis au sein de l'association. A fréquence fixe (mensuel, trimestriel, annuel), elle a la charge de collecter, contrôler et diffuser les données produites par l'ensemble des services de l'association.

Le projet décisionnel a été lancé fin 2013 afin de répondre aux enjeux de développement de l'association :

- Croissance externe et interne par la construction et l'acquisition de nouveaux établissements
- Diversification des activités, par la mise en place de services complémentaires destinés au grand âge.
- Augmentation du nombre de salariés afin de prendre en charge l'augmentation du niveau de dépendance des personnes accueillies.

Le développement de l'association engendre une augmentation du volume de données à traiter ainsi que de nouveaux besoins en rapports qu'il sera difficile de concilier avec les méthodes de productions actuelles. Il a donc été jugé nécessaire par la direction de revoir le mode de production des rapports. Cette réflexion se base sur deux axes :

- Evolution des processus
- Evolution des outils

En tant que membre du service système informatique rattaché à la direction financière, j'ai donc été chargé de mener ce projet en collaboration avec le contrôleur de gestion.

Après plusieurs mois d'étude et de préparation, en 2015 deux projets ont donc été lancés au sein d'EDENIS, le nouveau projet associatif destiné à redéfinir l'organisation humaine de l'association et le schéma directeur des systèmes d'information (SDSI) définissant une feuille de route sur trois ans (2015 -2018) de l'évolution du système d'information.

L'étude et la conception du SDSI a été faite avec l'appui du cabinet CAP GEMINI, il définit 19 projets et a notamment validé la poursuite et le développement du projet décisionnel désormais nommé en interne : projet « P4 Mise en place d'un Data warehouse (Base Reporting) ».

## I.3 Présentation du service système d'information d'EDENIS

### I.3.1 Organisation humaine

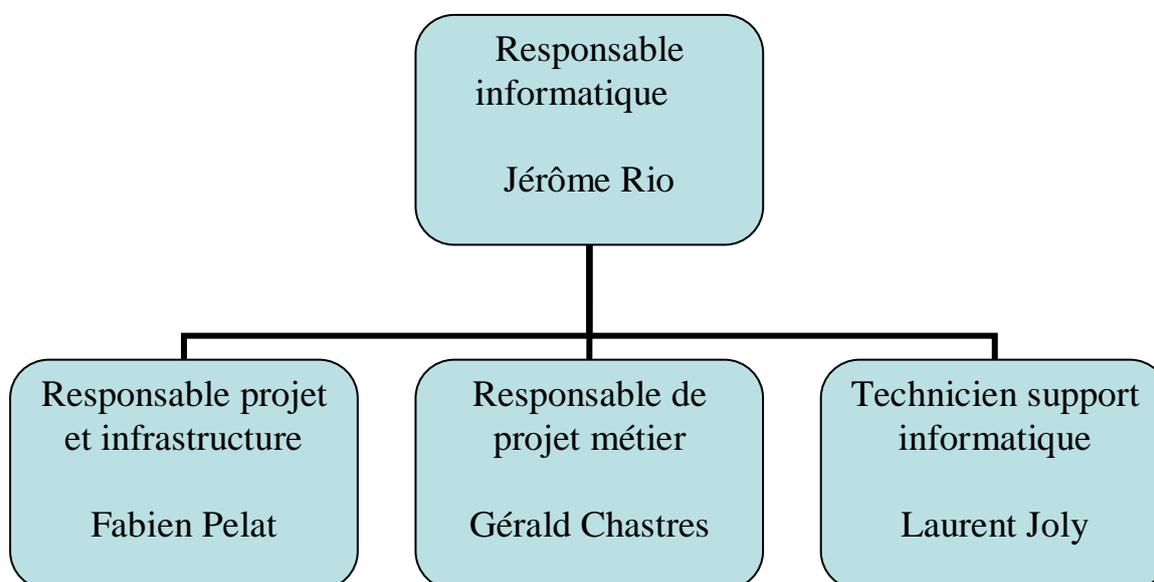
Le service informatique est composé de quatre personnes (Figure 3), suite au départ du précédent responsable informatique, M. Jérôme Rio a pris en janvier 2015 la direction du service en plus de son poste de contrôleur de gestion.

Le responsable informatique précédent avait en charge l'administration de l'infrastructure réseaux et serveurs et les projets informatiques ; rôle que j'ai repris à son départ.

Ma fonction de responsable des projets informatiques pour la direction financière a été transférée à M. Gérald Chastres qui assure désormais les fonctions de responsables projets informatiques pour l'ensemble du périmètre fonctionnel d'EDENIS.

M. Laurent Joly, technicien support est en charge de la gestion des matériels et des interventions sur site.

Le support informatique est assuré conjointement par l'ensemble des acteurs selon leurs domaines d'expertise.



### I.3.2 Cartographie Système d'information

L'infrastructure du système d'information d'EDENIS représentée dans la figure 4 est basée sur un réseau de type MPLS assurant l'interconnexion des différents sites de l'association.

La topologie du réseau MPLS a été revue au premier trimestre 2016, l'accès internet précédemment centralisé au siège a été déporté en cœur de réseau afin d'éliminer un point individuel de défaillance.

Les applications métiers sont hébergées dans une salle serveur climatisée, ondulée et protégée située au siège de l'association.

Cette salle héberge un cluster de virtualisation CITRIX d'ancienne génération et un cluster HYPER-V comprenant une ferme RDS 2012 Server mis en œuvre en janvier 2016.

Ce nouveau cluster permet de résoudre les problèmes d'obsolescence matériel et logiciel du cluster précédant et ainsi mettre à niveau la sécurisation de l'hébergement.

Les applications métiers y sont progressivement migrées depuis Janvier 2016.

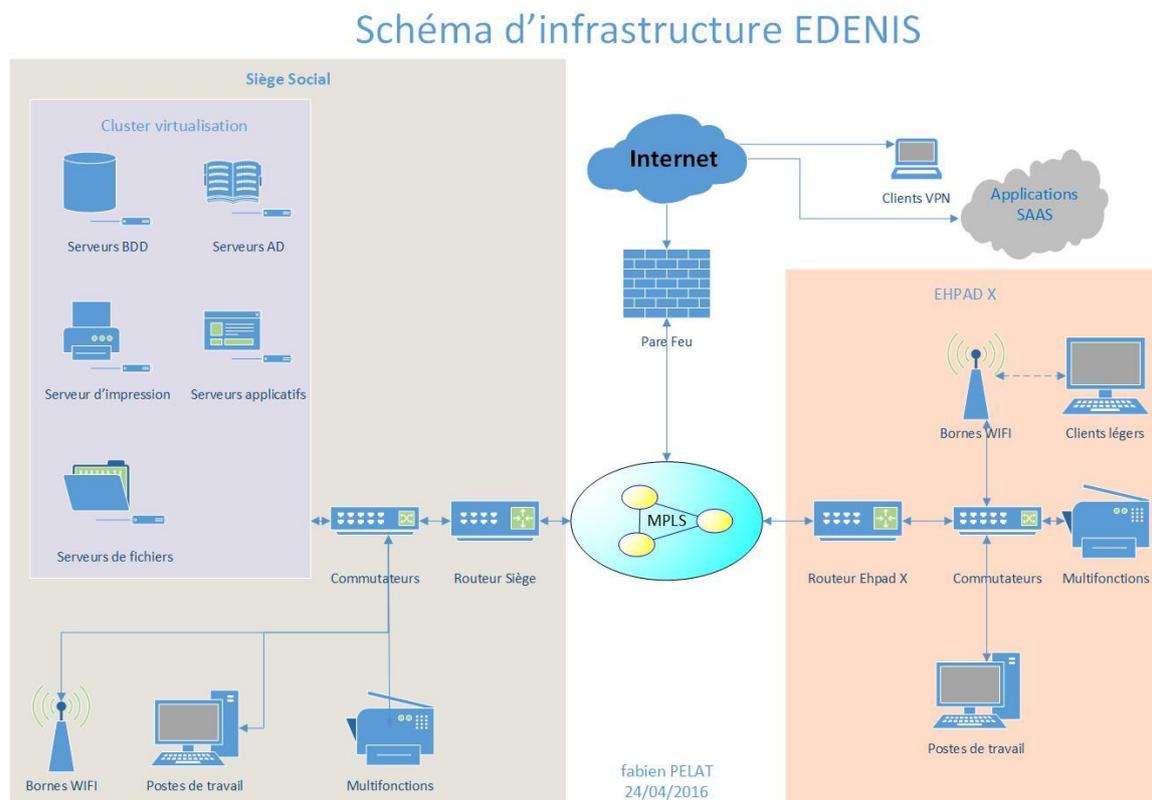


Figure 4 Schéma infrastructure EDENIS

Le système d'information d'EDENIS repose sur un ensemble de logiciels spécialisés (figure 5) pour chaque processus de l'association. Ces logiciels proviennent d'éditeurs et utilisent des technologies différentes (WEB / logiciel client ; FOX PRO / SQL). La communication entre elles s'effectue au travers d'échanges de fichiers manuels ou automatisés.

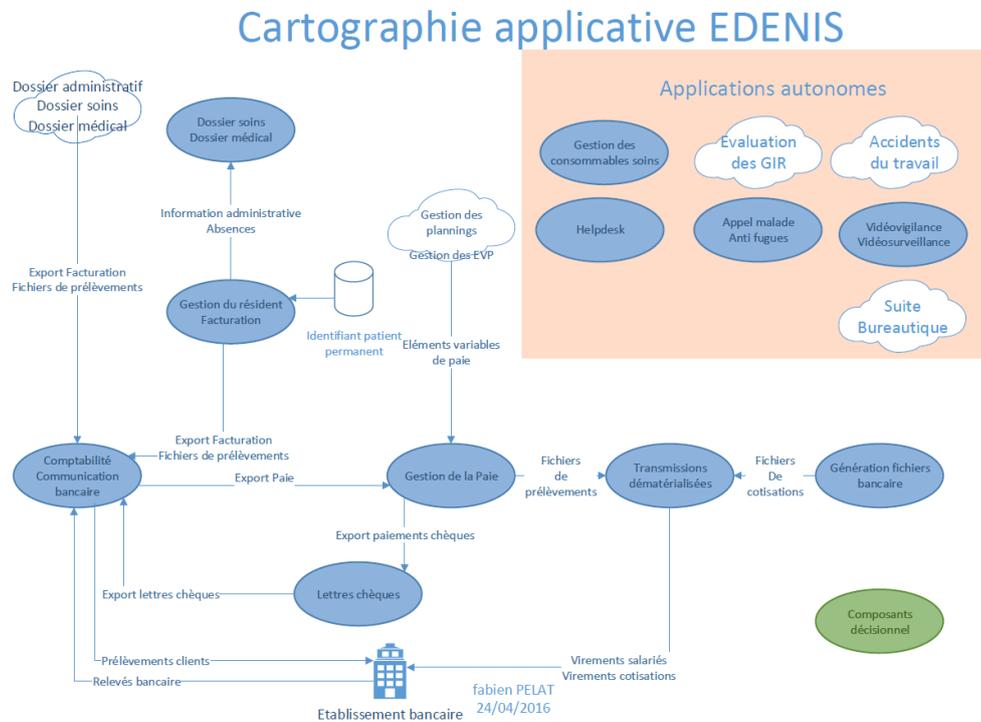


Figure 5 Cartographie applicative

Le schéma directeur des systèmes d'information livré au premier trimestre 2015 a identifié des domaines fonctionnels comme étant critique (Figure 6). Ces domaines sont en cours de migration vers des applications de type SAAS afin de garantir une haute disponibilité aux utilisateurs.

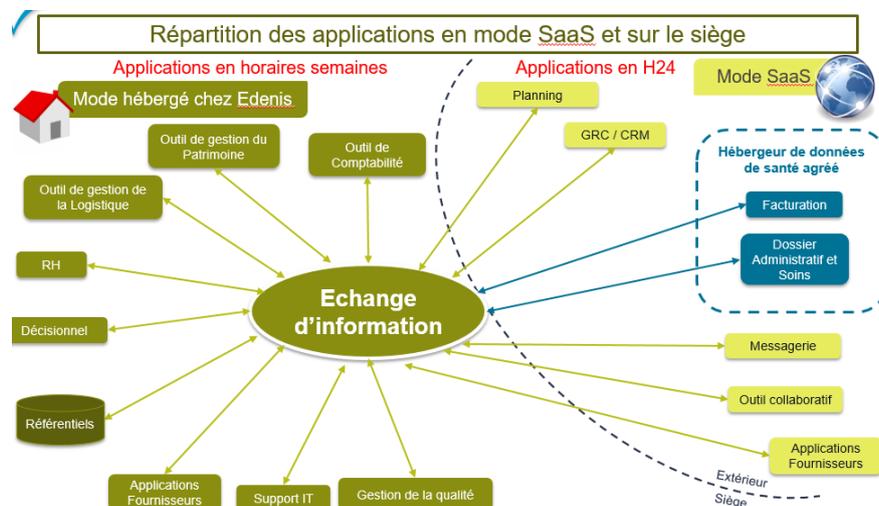


Figure 6 Répartition hébergement SDSI

## **II Etat de l'art**

### **II.1 Les origines**

### **II.2 Les besoins en pilotage**

Dans un contexte de crise et de concurrence accrue, piloter son entreprise efficacement est devenu indispensable pour en assurer la pérennité.

Fondamentalement, le pilotage d'une entreprise passe par sa capacité à faire le bilan de ses activités, processus par processus et à utiliser les données de ce bilan pour se projeter vers l'avenir sur la base d'objectifs et d'indicateurs de performance.

Deux outils principaux sont à la disposition du chef d'entreprise et des responsables opérationnels : le Reporting et les tableaux de bords.

Le Reporting dresse le bilan des activités, sous forme de données brutes, non traitées (le nombre de ventes, le nombre de ventes par produit, par commercial, la marge par produit, la marge rapportée au chiffre d'affaire général, le coût de revient, etc.) donc non pertinentes sans une analyse et une mise en perspective.

Les tableaux de bord utilisent les données fournies par le Reporting pour dresser une analyse de l'état de l'entreprise, de ces performances et établir des préconisations, fixer des objectifs d'amélioration, .... Les conditions de réalisation de ces objectifs seront décomposées de manière à pouvoir attribuer à chaque service en particulier un objectif précis qui participe à la réalisation de l'objectif global, et ainsi de suite....

Dans un système d'information d'entreprise, les données représentent un volume très important, dispersées dans les différentes bases de données ou sources de données réparties entre les différentes fonctions métiers.

Les applications transactionnelles ont été incapables de répondre de manière satisfaisante lorsque les entreprises eurent le besoin de réaliser une des activités suivantes :

- Analyse de données provenant de sources multiples
- Analyse de l'évolution historique de données
- Agrégation de quantité importantes de données

Les SGBD métier hébergent des bases de données relationnelles fonctionnant en mode transactionnel (OLTP On-Line Transaction Processing). Leurs objectifs sont de réaliser efficacement les opérations de sélection, ajout, mises à jour et suppressions de tuples, ces opérations étant réalisées pour de nombreux utilisateurs simultanés.

Le modèle relationnel de par sa conception ne peut répondre à des requêtes complexes sans dégradation forte des performances. Aucun mécanisme d'historisation des données n'est implémenté et ces bases de données étant souvent conçus par des éditeurs différents elles ne reposent pas sur des référentiels communs.

La conception d'un autre modèle de stockage des données était donc nécessaire, il s'agit des bases de données OLAP (On-Line Analytical Processing) définies par Bill Inmon (Inmon, 1990) de la manière suivante : « Un entrepôt de données (Data warehouse) est une collection de données thématiques, intégrées, non volatiles et historisées pour la prise de décisions »

- Thématiques : les données sont organisées par thème ayant un sens pour l'entreprise contrairement aux systèmes de production où les données sont organisées par processus fonctionnel.
- Intégrées : L'intégration dans le Data warehouse consiste à résoudre les problèmes d'hétérogénéité des données en s'assurant du respect de leur cohérence, normalisation, sémantique, contraintes et règles de gestion.
- Historisées et non volatiles : L'historisation consiste à suivre dans le temps l'évolution des différentes valeurs des données. Une fois les données stockées au niveau d'un Data Warehouse, les opérations de mise à jour ou de suppression ne sont plus autorisées. L'accès est autorisé uniquement en mode lecture.
- Prise de décisions : Les utilisateurs doivent avoir accès aux données détaillées, agrégées ou historisées qui leur sont autorisées dans l'entrepôts de données. Elles constituent la source d'alimentation des outils d'aide à la décision (rapports, OLAP, Dataminings...)

Ces bases de données ne seront utilisées par les utilisateurs qu'en mode lecture. Elles ont pour objectif de regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour donner à l'utilisateur une vue orientée métier, retrouver et analyser l'information facilement et rapidement.

Tableau 1 Comparatif OLTP/OLAP

	Online Transaction Processing (système métier)	Online Analytical Processing (Système décisionnel)
Source des données	Saisie et traitement effectués par les utilisateurs	Base de données OLTP et autres sources de données métier
Objectif des données	Réalisation et contrôle des traitements métier	Analyse et aide à la décision
Contenu des données	Image de l'activité métier actuelle et détaillée	Vue historisées et multidimensionnelle des activités métier
Mise à jour	En temps réel initiée par les utilisateurs	Traitement automatique planifié de rafraîchissement des données
Requête	Standardisées et simple retournant un nombre limité de tuples	Requêtes complexes comprenant des agrégations
Volume	Faible (Go)	Pouvant être très important du fait de l'historisation des données et des agrégations (To)
Temps de traitements	De l'ordre de la seconde	Pouvant atteindre plusieurs heures du fait de la complexité des requêtes et du volume des données
Modèle de base de données	Très normalisé avec de nombreuses tables	Peu de tables en étoiles ou en flocons

## II.3 Les théoriciens

Les principales dates à retenir construisant l'histoire de l'entrepôt de données sont les suivantes :

- Années 1960 - General Mills et l'Université Dartmouth, dans un projet conjoint, créent les termes *faits* et *dimensions*.
- 1983 - Teradata introduit dans sa base de données managériale un système exclusivement destiné à la prise de décision.
- 1988 - Barry Devlin et Paul Murphy publient l'article *Une architecture pour les systèmes d'information financiers (An architecture for a business and information systems)* où ils utilisent pour la première fois le terme *Data warehouse*.
- 1989, Howard Dresner, un chercheur du Gartner Groupe, a popularisé le terme « Business Intelligence », équivalent anglophone de l'informatique décisionnelle. Il l'a décrit comme étant un ensemble de concepts et de méthodes pour améliorer la prise de décisions en utilisant des systèmes s'appuyant sur des données concrètes et réelles.
- 1990 - Red Brick Systems crée Red Brick Warehouse, un système spécifiquement dédié à la construction de l'entrepôt de données.
- 1991 - Bill Inmon publie *Building the Data Warehouse (Construire l'entrepôt de données)*.
- 1995 - Le *Data Warehousing Institute*, une organisation à but lucratif destinée à promouvoir le data warehousing, est fondé.
- 1996 - Ralph Kimball publie *The Data Warehouse Toolkit (La boîte à outils de l'entrepôt de données)*.

## II.4 Concepts Décisionnel

### II.4.1 Modélisation des informations d'aide à la décision

La modélisation de base de données OLAP repose sur le concept de tables de fait et de tables de dimension

### II.4.2 Concept de tables de fait.

#### **Définition R. KIMBALL d'une table de fait :**

Une table de fait est une table qui contient les valeurs numériques produites par un évènement opérationnel mesurable dans la vie réelle. Au niveau le plus fin, un enregistrement d'une table de fait correspondant à la mesure d'un évènement et inversement. Ainsi le design fondamental d'une table de fait est entièrement basé sur l'activité réelle et ne doit pas être influencé par les rapports qui devront être produits. (KIMBALL et ROSS 2016)

Les types de table de fait :

- Table de fait transactionnelle :

Il s'agit du type de table le plus commun, elle représente les transactions d'un système opérationnel, sa granularité peut être déterminée par le niveau d'agrégation des données.

- Table de fait périodique :

Elle représente une mesure régulière d'un fait, l'horodatage représente la fin d'une période cohérente avec le fait mesure (entrée/sortie du jours, solde mensuelle, résultat annuel). Elles permettent d'analyser le résultat d'un volume de données très important sur de longues périodes

- Table de fait récapitulative :

Une table de fait récapitulative mesure des actions métier sur une période non prédéterminée elle intègre typiquement de nombreuses dimensions de type date permettant de suivre le cycle de vie d'un fait. Il s'agit du seul type de fait pour lequel on effectue des opérations de type mise à jour.

Tableau 1 Comparatif caractéristiques types tables de fait

<b>Table de fait</b>	<b>Transactionnelle</b>	<b>Périodique</b>	<b>Récapitulative</b>
Granularité	Une ligne par transaction	Une ligne par période	Une ligne pour toute la durée de vie d'un événement
Dimension	Dimension de date au niveau de granularité le plus bas	Dimension de date au niveau de granularité Fin de période	Plusieurs dimensions de date
Nombre de dimensions	Plus élevé que le type de fait périodique	Moins élevé que le type de fait transactionnel	Nombre de dimensions le plus élevé par rapport aux autres types de table de fait
Mesures	Relatives aux activités de transaction	Relatives aux activités périodiques	Relatives aux activités qui possèdent une durée de vie définie
Taille de la base de données	Taille importante.	Réduite par rapport à la table de fait transactionnelle en fonction de la granularité de la dimension date	Faible
Performances	Offre de bonnes performances qui peuvent être améliorées en sélectionnant un grain situé au-dessus du grain le plus détaillé	Offre de meilleures performances que les autres types de table de fait car des données sont stockées au niveau d'un grain moins détaillé	Bonnes performances
Insertion	Oui	Oui	Oui
Mise à jour	Non	Non	Oui
Suppression	Non	Non	Non
Augmentation de la taille de la table des fait	Importante	Moyenne	Faible

### II.4.3 Concept de dimensions

#### **Définition R. KIMBALL d'une table de dimension :**

Chaque table de dimension a une unique colonne de clef primaire. Cette clef primaire est intégrée comme clef étrangère dans toutes les tables de fait associées ou le contexte de la ligne de dimension correspond à la ligne de la table de fait. Les tables de dimension sont en général de grande tables dé normalisées. (KIMBALL et ROSS 2016)

Les types de dimension :

- Dimension dégénérée :

Une dimension dégénérée consiste à ajouter dans une table de fait une colonne faisant référence à un identifiant de la base de données transactionnelles source.

- Dimension à évolution rapide

Une dimension à changement rapide est une dimension qui subit des changements très fréquents des attributs dont on veut préserver l'historique.

- Dimension conforme :

Une dimension conforme est une dimension utilisée par plusieurs faits ou des dimensions partageant les mêmes attributs. Ces tables de fait peuvent être combinées dans un seul rapport en utilisant les attributs de dimension conforme. Les dimensions conformes, définies une fois en collaboration pour toute l'entreprise, sont réutilisés dans les tables de fait ; elles offrent à la fois une cohérence pour l'analyse et facilitent les futurs développements

- Mini dimension

Une mini dimension est un sous ensemble d'attribut, changeant rapidement, d'une grande dimension, causant l'augmentation excessive de la dimension dans le cas de l'utilisation de la technique de dimension à évolution lente de type 2. Les valeurs des attributs sont extraites dans une mini dimension reliée directement à la table de fait.

- Dimension à évolution lente :

R. KIMBALL identifie 7 sept types de dimensions à évolution lente, les types 1 , 2,3 et 4 sont les plus utilisés. (KIMBALL et ROSS 2016)

Tableau 2 Sept types de dimensions à évolution lente

Type	Action sur la table de dimension	Impact sur l'analyse des faits
Type 0 : Original	Pas de changement de la valeur de l'attribut	Les faits sont associés à la valeur originale de l'attribut
Type 1 : remplacement	Remplacement de la valeur de l'attribut	Les faits sont associés à la valeur actuelle de l'attribut
Type 2 : nouvelle ligne	Ajout d'une nouvelle ligne d'attribut pour intégrer la nouvelle valeur	Le fait est associé à la valeur de l'attribut en cours lors de l'évènement
Type 3 : Nouveau attribut	Ajout d'une nouvelle colonne pour conserver l'attribut existant et prioriser les valeurs	Les faits sont associés à la fois avec la valeur actuelle et la valeur prioritaire de l'attribut
Type 4 Historisation	Ajout d'une table de mini dimension contenant les valeurs à changement rapide	Le fait est associé à la valeur de l'attribut à changement rapide en cours lors de l'évènement
Type 5 hybride	Ajout d'une mini dimension de type 4 associé au remplacement de type 1 de la clef de mini dimension	Le fait est associé à la valeur de l'attribut à changement rapide en cours lors de l'évènement et aux valeurs de l'attribut à changement rapide
Type 6 hybride	Ajout d'attributs de type 1 aux lignes de type 2	Les faits sont associés aux valeurs d'attribut en cours au moment de l'évènement ainsi qu'aux valeurs actuelles.
Type 7 hybride	Ajout de ligne de type 2 avec de nouvelles valeurs d'attributs, limité aux au valeurs de lignes actuelles	Les faits sont associés aux valeurs d'attribut en cours au moment de l'évènement ainsi qu'aux valeurs actuelles.

#### **II.4.4 La modélisation en étoile**

La modélisation en étoile, initiée par Ralph Kimball, consiste en la création d'une table de fait à laquelle sont associées des tables de dimension. Chaque dimension est décrite par une seule table dont les attributs peuvent représenter toutes les granularités possibles.

Cette modélisation ne respecte pas les formes normales généralement associées à la modélisation dimensionnelle afin de privilégier les performances de restitution des données au détriment du volume de stockage.

#### **II.4.5 La modélisation en flocon**

Dans un schéma en flocon, initié par Inmon, la différence avec la modélisation en étoile se situe au niveau des tables de dimension qui peuvent être organisées en hiérarchie. Les tables de dimension sont reliées entre elles pour représenter la granularité de l'information

La hiérarchisation des dimensions permet d'éviter la redondance d'information induite par la modélisation en étoile mais nécessite des jointures lors des agrégats de ces dimensions via une succession de clés étrangères.

R. Kimball (KIMBALL et REEVES, Concevoir et déployer un data warehouse 2000) a évalué que la modélisation en flocon permet un gain de place disque d'environ 1 %, il juge que ce gain ne compense pas les deux inconvénients suivants :

- la normalisation ajoute des jointures qui ralentissent l'exécution des requêtes.
- le modèle est plus difficile à appréhender par les utilisateurs non informaticiens

#### **II.4.6 La modélisation en constellation**

La modélisation en constellation de faits est le résultat de la modélisation de plusieurs tables de fait partageant des tables de dimension conformes.

## **II.5 Architecture d'un système décisionnel**

Un système décisionnel se décompose traditionnellement en 4 composants :

### **II.5.1 Composant de collecte**

Un système d'information d'entreprise est par nature hétérogène sur le plan logique et technique. Les sources de données sont nombreuses, variées, distribuées et autonomes, elles peuvent être internes (bases de données de production, ERP, Archives, Feuilles de calcul ou externes (Internet, bases des partenaires) à l'entreprise (TESTE, 2000). Avant d'être utilisable les données doivent être extraites de leurs sources puis être uniformisées, nettoyées et consolidées avant d'être chargées dans le système décisionnel. Ces rôles sont assurés par les outils d'ETL (Extract, Transform and Load)

L'architecture ETL requière une multitude de sous-systèmes nécessaires pour répondre aux exigences imposées à l'entrepôt de données. Ralph Kimball, (KIMBALL, Intelligent Enterprise, 2004) a défini 34 sous-systèmes composant un ETL et a estimé que 70% du temps et des efforts nécessaires à la mise en œuvre d'un projet d'entrepôt de données est dédié aux systèmes ETL.

Etape 1 : Extraction des données :

La première phase d'un ETL consiste à accéder aux données de production selon plusieurs moyens selon la nature de la source de données et l'impact que le processus peut causer sur ses performances :

- Utilisation de connecteurs sur les données de production
- Réplication des données de production dans l'environnement décisionnel
- Planification d'export de données à partir des outils métiers

Etape 2 : Transformation des données :

Il s'agit de l'étape la plus critique de la chaîne décisionnelle car elle réconcilie les données entre les différentes sources. A cette étape, les données peuvent être vérifiées, agrégées, calculées, dédoublonnées, enrichies avec des données externes avant d'être structurées pour respecter le format requis par les systèmes cibles (Troisième Forme Normale, Schéma en Etoile, Dimensions à Evolution Lente, etc.)

Etape 3 : Le chargement de données :

Cette étape charge dans le composant de stockage les données résultantes des étapes précédentes.

### **II.5.2 Composant de stockage**

Comme nous l'avons vu dans les besoins à l'origine de la création des systèmes décisionnelles, la modélisation décisionnelle est conçue pour répondre à des objectifs très différents des bases de données relationnelles. Les données sont donc stockées dans des bases de données spécialisées qui peuvent être selon les besoins :

- L'ODS (Opérationnal Data Store) :
- Le Data warehouse :
- Le Datamart :

### **II.5.3 Composant de diffusion**

De par sa transversalité et le volume de données gérées il n'est pas souhaitable que les utilisateurs aient un accès direct à l'entrepôt de données. Il est nécessaire de gérer les droits d'accès aux données ainsi que dès la mise à disposition de schémas thématiques. Le composant de diffusion met ainsi les données à la disposition des utilisateurs, selon les droits et les schémas correspondant à leurs profils. Le composant de diffusion se présente sous la forme d'un portail WEB de business intelligence intégré au portail d'entreprise

### **II.5.4 Composant d'exploitation**

Une fois les données stockées, nettoyées, consolidées et accessibles, elles sont utilisables. Selon les besoins, différents types d'outils d'exploitation seront envisagés.

- Le besoin consiste à analyser des volumes important de données, il convient d'envisager des outils de type OLAP (Cube, hyper cube, ...) pour les analyses multidimensionnelles.
- Le besoin consiste à rechercher des corrélations peu visibles, il convient d'envisager des outils de Datamining .
- Le besoin est orienté pilotage et aide à la décision, il convient d'envisager des outils de création de tableaux de bord présentant les indicateurs clés de l'activité.
- Le besoin est orienté diffusion des données, il convient d'envisager des outils de Reporting.

## **II.6 Mise en œuvre d'un projet décisionnel**

### **II.6.1 Les préalables projet.**

. Selon Kimball (KIMBALL, ROSS, Thornthwaite, MUNDY, & BECKER, 2008), il y a trois points essentiels que les organisations doivent évaluer avant de se préparer à faire un projet BI :

- Le niveau d'engagement et le sponsoring du projet par la direction.
- Le niveau de demande utilisateur pour la mise en œuvre d'un projet BI.
- La quantité et la qualité des données d'entreprise disponibles

### **II.6.2 Le sponsoring**

Comme le font remarquer Kimball et al (KIMBALL, ROSS, Thornthwaite, MUNDY, & BECKER, 2008) ,même le système Data warehouse / Business Intelligence le plus élégamment conçu ne peut pas surmonter un manque de soutien de la direction.

Un projet décisionnel est par nature transversal au sein de l'organisation, il pourra donc être impacté par des blocages ou des difficultés issues de tous les services composant l'entreprise. Seul un parrainage de haut niveau pourra permettre de lever ou d'arbitrer ces blocages.

### **II.6.3 L'intérêt métier.**

Un projet décisionnel trouve sa justification auprès des métiers dans les cas de figures suivants :

- Le gain d'un avantage commercial
- L'acquisition de nouvelles activités hors du cœur de métier d'origine de l'entreprise
- La gestion de multiples filiales
- Un manque de disponibilité et de gestion des données clefs de l'entreprise

## II.6.4 La quantité et la qualité des données

Avant la mise en œuvre de tout projet décisionnel il est important de faire du profilage des données. Cette analyse identifie le contenu, la cohérence et de la structure des données afin de déterminer si des données manquent et la manière de les récupérer.

Pour chaque domaine métier les besoins sont généralement les suivants :

1. Inventorier les sources de données afin de répertorier les données nécessaires
2. Convertir les données métiers en informations cohérentes
3. Requêter et analyser les données
4. Décider à partir des données collectées

Dans un projet décisionnel, la gestion de la qualité des données devrait être un processus de bout en bout, à savoir de la source de données au rapport final.

Etapes	Tâches
Sources de données	<ul style="list-style-type: none"><li>• Standardisation des données (unités, modèles, ...)</li><li>• Référentiel de données unique.</li></ul>
Opérationnal data store	<ul style="list-style-type: none"><li>• Nettoyage des données (détecter et corriger les données métiers inexacts)</li><li>• Profilage des données (nettoyer les données valeur Nulles, vides, ...)</li></ul>
Data warehouse	<ul style="list-style-type: none"><li>• Contrôle du chargement intégral des données</li><li>• Référentiel unique de toutes les sources de données</li><li>• Contrôle de cohérence entres les sources</li></ul>
Reporting	<ul style="list-style-type: none"><li>• Dictionnaire des indicateurs</li><li>• Contrôle et validation des formules de calculs</li></ul>

Figure 7 Gestion de la qualité des données

## II.6.5 La sélection des domaines à traiter

Il peut être difficile d'identifier un domaine métier pertinent pour démarrer un projet de business intelligence. La priorisation des projets BI dans une entreprise peut se faire en considérant les points suivants :

- La facilitation de l'accès aux données d'un domaine métier pour l'ensemble des employés permet un gain global important pour l'entreprise même si chaque employé ne gagne que quelques minutes
- R. KIMBALL a indiqué en 2008 (KIMBALL, ROSS, Thornthwaite, MUNDY, & BECKER, 2008) le responsable de projet BI doit identifier les bénéfices concrets du projet décisionnel comme par exemple l'élimination des coûts actuels de production des rapports
- Mettre en place une matrice de décision afin de déterminer une méthodologie cohérente avec les besoins de l'entreprise.

## II.6.6 Le cycle de vie d'un projet décisionnel

Généralement, le développement d'un système décisionnel se fait en suivant l'une de ces deux méthodes : celle de Kimball ou celle d'Inmon (GOEDE & HUISMAN, 2010)

### II.6.6.1 Méthode KIMBALL

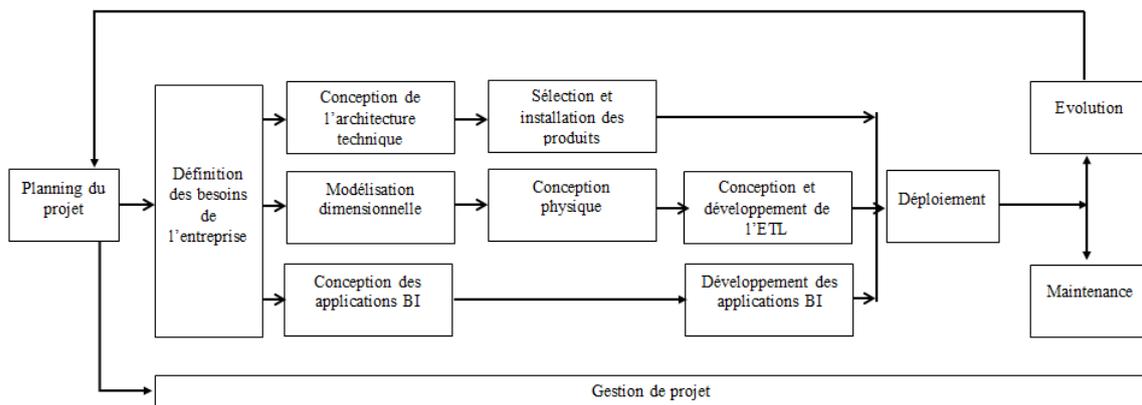


Figure 8 Le cycle de vie d'un projet décisionnel (KIMBALL & ROSS, THE Data Warehouse Toolkit : The Definitive guide to dimensional modeling 3rd Edition, 2013)

La méthode de Kimball débute par les besoins du client. Par la suite, la création de l'entrepôt de données, tels que trois chemins sont empruntés en parallèles, car ils ne visent pas les mêmes éléments de l'environnement (RIVARD)

- La conception de l'architecture technique : □ on choisit l'architecture technique et les outils matériels et logiciels nécessaires pour la mise en place de l'entrepôt de données.
- La modélisation dimensionnelle : on définit la modélisation dimensionnelle de l'entrepôt et des magasins de données, et on définit les outils ETL.
- Conception des applications BI : développement des applications décisionnelles tels que : les rapports, les tableaux de bord...

Ces trois chemins convergent en fin du projet au moment du déploiement. Le processus complet est répété pour chaque nouveau magasin de données demandé par les utilisateurs finaux tout en assurant l'évolution et la maintenabilité du système. (RIVARD)

## II.6.6.2 Méthode INMON

La méthode d'Inmon est basée sur les points suivants (INMON W. , 2002)

- La création de l'entrepôt de données, ce dernier permettra d'alimenter les autres magasins de données des différents départements ce qui garantit la cohérence de données.
- Une connaissance approfondie du domaine métier de l'entreprise.
- La modélisation de données repose sur le modèle Entité/Association et la normalisation en 3<sup>ème</sup> forme normale.
- Inmon divise la base de données de l'entreprise en quatre niveaux :
  - Opérationnel : contient les BDD du système opérationnel ;
  - Entrepôt de données atomiques ;
  - Départemental ;
  - Individuel.

Les trois derniers niveaux constituent l'entrepôt de données. Les données sont transformées entre le premier et le deuxième niveau en utilisant un processus ETL.

Cette méthode est une approche menée par les données.

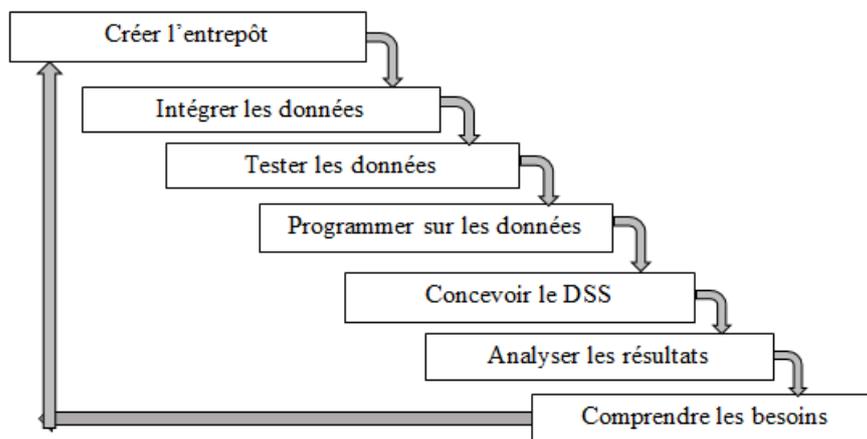


Figure 9 Le cycle de vie d'un projet décisionnel selon la méthode Inmon (INMON W. , 2002)

Selon le schéma ci-dessus, la création de l'entrepôt est la première étape du projet, et selon les résultats de la conception retenus à la fin du cycle, les besoins définis par le client seront bien compris par l'équipe projet ce qui permet de reprendre le cycle du projet pour permettre l'évolution de l'entrepôt de données et non pas sa révolution

### II.6.6.3 Comparaison entre les deux méthodes

Les méthodes d’Inmon et Kimball sont différentes dans leur globalité, mais elles ont les mêmes caractéristiques et suivent le même type de cycle.

Le tableau ci-dessous, synthétise les différences majeures entre les deux méthodes, en se basant sur les travaux de recherche de Breslin. (BRESLIN, 2004)

Méthode	Inmon	Kimball
Approche	Top-Down	Bottom-Up
Structure	Le Data warehouse alimente les Data Marts	Les Data Marts modélisent un processus métier, le schéma de l’entreprise est donné par le bus de données et les dimensions conformes
Complexité	Complexe	Simple
Processus	Dérive de la méthode spirale	Processus à quatre étapes
Architecture	Complexe	Simple
Outil	Traditionnels	Modélisation dimensionnelle
Accessibilité de l’utilisateur	Faible	Forte
Audience	Professionnels	Utilisateurs finaux

**Tableau 1:** Comparaison entre les approches Kimball et Inmon (BRESLIN, 2004)

## II.7 Le marché du décisionnel

### II.7.1 Historique

Tableau 3 Historique marché décisionnel

Année	Evénement
1962	Ken Iverson publie le langage APL, le premier langage multidimensionnel
1970	Express est le premier outil multidimensionnel destiné aux applications marketing
1982	Comshare System W premier outil OLAP à destination des applications financières
1983	Teradata introduit dans sa base de données managériale un système exclusivement destiné à la prise de décision.
1984	Métaphor le premier moteur ROLAP
1990	Cognos powerplay premier client OLAP pour station de travail
1993	E.CODD dicte les règles qui décrivent les moteurs OLAP
1998	Microsoft lance OLE DB for OLAP un ETL et un moteur multidimensionnel.
2000	Microsoft intègre gratuitement Analysis Services à SQL server 2000
2003	BO achète Crystal Decisions, Hyperion achète Brio Software, Cognos achète Adaytum et Geac achète Comshare
2004	BO, COGNOS, Microsoft, Microstratégie et Oracle mette à disposition des add-ins pour le tableur EXCEL
2005	Microsoft intègre à SQL serveur 2005 un studio de développement destiné au métier du décisionnel BI (data warehouse , TEL , OLAP , DATA Minig)

Les six plus grands éditeurs à savoir (Oracle, SAP, IBM, Microsoft, SAS, Teradata) représentaient en 2015 les deux tiers du marché, 64% pour être précis.

## II.7.2 Les tendances

Le marché français du décisionnel frôle les 10 % de croissance annuelle, et le taux d'adoption des solutions BI par les utilisateurs ne cessent d'augmenter. Cependant il s'agit d'un domaine qui continue d'évoluer fortement de nombreuses tendances émergent.

### 1) Technologies

#### a. Virtualisation

Il s'agit d'intégrer la BI dans le processus de virtualisation des infrastructures afin d'amplifier les économies d'échelles

#### b. Open source

Les solutions open source commerciales dévoilent des solutions innovantes qui prennent des parts de marché aux acteurs traditionnels (Talend, Jaspersoft,...)

#### c. Appliance

De plus en plus de grands acteurs (Oracle, Microsoft) proposent une architecture optimisée incluant le matériel, le logiciel et les services

#### d. SAAS/CLOUD

Souvent moins coûteuse et plus rapide à mettre en place, une solution Cloud peut être en mesure de réduire les délais de mise en place de quelques mois à quelques semaines seulement. Le marché de la location d'application BI est en forte augmentation avec une multiplication des offres malgré les craintes liées à la sécurité et à la confidentialité.

#### e. Intégration

Il s'agit de la mise en œuvre de solutions intégrées système décisionnel et système métier pour des secteurs d'activité ou des métiers

### 2) Données

#### a. Big Data

La croissance toujours plus rapide des volumes de données à analyser constitue un véritable challenge pour les plateformes décisionnelles traditionnelles qui doivent maintenant être capables de traiter des milliards de lignes de données de sources variées sans affecter leur temps de réponse. C'est possible grâce à des technologies « in memory » qui permettent de traiter des milliards de lignes de données en quelques secondes seulement.

#### b. Intégration de données non structurées

Les utilisateurs sont sources de nombreuses données non prises en compte par les solutions classiques textes (emails, fichiers), audio (VOIP), vidéo, contributions web...

De plus, de nouvelles méthodes de gestion de bases de données apparaissent comme les bases de données en colonnes (vectorielles) ou les bases de données multimédia.

La BI ayant vocation à intégrer toutes les sources de données, elle doit prendre en compte ces sources.

### 3) Fonctionnalité

#### a. Temps réel

Une autre tendance est le fait de permettre l'analyse des données en temps réel afin de pouvoir améliorer la réactivité des managers en cas de problème. Des fonctions d'alertes par exemple permettent aux utilisateurs de rester connectés à tout moment, de réagir afin de rectifier la tendance mais aussi d'anticiper pour l'avenir.

#### b. Interfaces 2.0 et collaboration

Mise à jour des interfaces de restitution afin d'apporter des fonctionnalités de conversations, commentaires, notations, partages et leurs intégrations aux réseaux sociaux d'entreprise.

#### c. Mobilité

Les managers sont de plus en plus mobiles, et les solutions de business intelligence « web » ou accessibles depuis un navigateur depuis n'importe quel appareil permettent de répondre pour les entreprises et les utilisateurs.

#### d. Solution prédictive

Le « prédictif » permet de savoir avant les autres, il s'agit donc de mettre en œuvre de nouvelles méthodes de prévision et de nouveaux types d'indicateurs pour anticiper et d'agir en avance de phase

#### e. Management des processus business

Il s'agit d'intégrer la BI au plus près des applications opérationnelles afin de fournir de l'aide à la décision aux opérateurs métier dans leurs actions quotidiennes.

## **III Projet décisionnel EDENIS**

### **III.1 Origines du projet**

#### **III.1.1 Les besoins à l'origine du projet**

L'association n'ayant pas identifié de solution de type ERP permettant la gestion de son activité, chaque processus métier est soutenu par des logiciels spécialisés (Facturation, Comptabilité, Ressource humaine, Gestion administratives du résident, ...). Edenis exploite un ensemble de progiciel possédant leurs propres moteurs de rapports intégrés.

Le contrôle de gestion et la direction financière ont exprimé auprès du service informatique les difficultés suivantes :

- La multitude de moteurs de rapports est difficile à maîtriser car ils n'utilisent pas de syntaxe identique ou peuvent être limités par l'éditeur.
- La consolidation de données utilisant des codifications propres à chaque outil est problématique car elle nécessite des tables de correspondances générées manuellement et donc source d'erreur.
- La perte d'information causée par l'absence d'historisation dans les logiciels métiers (un même rapport exécuté à un moment différent peut produire des résultats différents sans que cet écart soit justifiable car une donnée aura été supprimé par un opérateur de saisie)
- Le manque de pertinence des outils utilisés, l'utilisation d'un tableur n'est pas adaptée quand la volumétrie des données dépasse le million de ligne. Un tableur ne permet également pas de maîtriser efficacement la diffusion des données.

### **III.1.2 Historique des projets précédents.**

Les besoins à l'origine du projet sont exprimés depuis plusieurs années, sans que des solutions efficaces soient apportées.

En 2004 fut lancé un projet d'automatisation des rapports financiers basés sur des liens ODBC vers les bases de données comptables et des macros Excel. Ce projet a échoué au bout de plusieurs mois de développement car l'augmentation de la volumétrie des données causée par l'historisation et l'augmentation du périmètre couverts à rapidement conduit à des temps de traitement incompatible avec les exigences métiers. Il est apparu que la solution technologique n'aurait pas pu permettre de gérer l'intégralité des données de l'association sur plusieurs années.

En 2012 un autre projet fut lancé, à l'occasion du changement de logiciel de comptabilité. L'éditeur de celui-ci nous a proposé une solution de Reporting basé sur les fonctionnalités de business intelligence de Microsoft SQL SERVER. Cette solution intégrait un Data warehouse préconçu et ses ETL d'alimentation ainsi que la mise à disposition d'un portail rapport.

Les modalités contractuelles ne permettaient pas à EDENIS de réaliser la moindre modification de la solution hormis la personnalisation des rapports. Tout ajout ou modification de données au Data warehouse était interdit à moins de les faire réaliser par l'éditeur au travers de développements spécifiques.

Les tests menés en interne ont validé la capacité de la solution à répondre à nos besoins pour la partie Reporting comptable mais la direction financière n'a pas jugé pertinent de faire réaliser des développements spécifiques sur l'ensemble de nos données par un éditeur ne maîtrisant pas notre cœur de métier.

Ce projet a été stoppé mais son bilan a permis de tirer les conclusions suivantes :

- La solution de business intelligence de Microsoft SQL SERVER peut répondre aux besoins exprimés.
- Ce type de projet nécessite une connaissance approfondie des processus métier de l'association
- Ce projet évoluera en parallèle du système d'information, il est donc nécessaire de disposer de compétences disponibles et réactives afin de traiter toutes les modifications nécessaires

### III.1.3 Pré-étude du nouveau projet.

J'ai été fortement impliqué lors des tests des deux projets précédents et de par mes fonctions au service informatique, je réalise très régulièrement des rapports ou des extractions de données à partir des outils de Reporting intégrés dans les logiciels existant. J'ai donc perçus une opportunité de proposer une solution pérenne répondant aux besoins de l'association.

Néanmoins avant de proposer le projet à la direction, j'ai réalisé des études afin d'en valider la pertinence.

- Analyse des besoins fonctionnel Afin de formaliser les besoins exprimés j'ai collecté les besoins utilisateurs exprimés lors des différents projets ainsi que par les demandes de services soumises auprès du service informatique. J'ai également posé des questions de type « interview » lors de mes rencontres avec les différents responsables de processus.

A l'issue de cette collecte, les exigences suivantes ont été formalisées pour le projet :

1. L'objectif du projet est de résoudre les difficultés de Reporting des applications métier sans impacter l'existant.
2. La solution devra permettre au comité de direction et aux responsables d'accéder de manière autonome à une bibliothèque de rapports
3. Les autres utilisateurs devront pouvoir recevoir des rapports par mail.
4. Les données gérées dans la solution devront être codifiées selon un référentiel propre à EDENIS.
5. La solution devra permettre la création de rapports contenant des données consolidées provenant de plusieurs sources métiers (gestion du résident / comptabilité, planning / paie)
6. Le projet doit prendre en compte le besoin pour certaines données, d'historiser leurs évolutions. (Changement annuel d'un tarif par exemple) même après la suppression de ces informations dans les bases de données de production ou de l'évolution du système d'information afin de permettre une continuité et une fiabilisation de la production des rapports
7. Les traitements automatisés devront être optimisés pour permettre la production de rapports dans des délais correspondant aux besoins métiers. Pour chaque traitement, un plan de mise à jour sera défini avec les utilisateurs, il déterminera les fréquences

de mises à jour, les horaires de démarrage, la durée maximale du traitement ainsi que les modalités de déclenchement.

8. La solution devra permettre l'intégration de données provenant d'établissements tiers possédant leurs propres systèmes d'information mais rattaché à EDENIS dans le cadre de sa croissance externe (reprise en gestion ou intégration)

Les exigences ont été traduites en fonctionnalités informatiques permettant de déterminer les caractéristiques de la future solution de de vérifier leurs disponibilités dans l'outil proposé.

Tableau 4 liste des fonctionnalités attendues

Fonctionnalité	Type de composants
Accès aux données EDENIS	Intégration de données
Gestion de tous les formats de données	Intégration de données
Accès aux sources de données tierces	Intégration de données
Planification des actualisations des données	Intégration de données
Centralisation des données	Stockage
Historisation, agrégation, consolidation des données	Stockage
Création et consultation de rapports	Diffusion
Dictionnaire des données	Documentation
Envoi de rapport	Diffusion
Bibliothèque de rapports	Diffusion
Outil d'analyse des données	Exploitation
Autonomie des utilisateurs pour la création de rapports	Exploitation

- Analyse des sources de données

L'analyse des besoins utilisateurs a notamment permis d'identifier de nombreuses sources de données et tables de correspondances sous formes de feuilles de calculs non gérées par le service informatique mais indispensables au projet.

Ces sources de données ont été ajoutés au recensement des bases de données par le service informatique afin de réaliser une étude de volumétrie et de déterminer tous les formats de sources de données devant être pris en compte par la future solution.

Tableau 5 Liste des formats de données gérés par l'association

Format de données	Volumétrie totale	Quantité
SQL	35 Go	5
Tableurs ( xls ;xlsx)	200 Mo	100
Fichiers (csv)	18 Go	18

- Analyse des solutions du marché.

Au moment du lancement du projet, le marché du décisionnel était composé de deux groupes, les acteurs de niches et les acteurs leaders. Je me suis donc intéressé aux acteurs leaders à savoir Oracle, Microsoft, IBM, SAS, Microstrategy, SAP , Information Builders. Aucune de ces solutions n'est open source et elles nécessitent toutes un investissement de licence important.

Cependant la solution approuvée lors du projet précédant fait bien partie des leaders du marché.

J'ai réalisé un inventaire de licence, il est apparu que l'association disposait d'une licence inutilisée de la suite Microsoft SQL Serveur Standard Edition permettant l'accès à l'ensemble des composants décisionnels.

La solution Microsoft ayant été testée et validée lors d'un précédent projet, elle répond à tous les besoins fonctionnels exprimés par les utilisateurs et les investissements financiers pour l'acquérir ont déjà été réalisés. J'ai donc naturellement basé ma proposition de projet sur cette solution.

### III.1.4 Proposition projet

L'analyse des besoins et des sources de données m'a permis d'affiner ma proposition de projet et ainsi de la soumettre à la direction financière. J'ai donc proposé la réalisation d'un entrepôt de données basé sur la solution Microsoft SQL serveur dont les composants seraient développés en interne au sein d'EDENIS. Le service informatique se chargerait des ETL et de l'entrepôt de données, le contrôle de gestion de la conception des rapports.

Ma proposition de projet est basée sur les arguments suivants :

- Un projet de ce type nécessite une connaissance approfondis des processus EDENIS
- La solution évoluera en parallèle du système d'information d'EDENIS, il est donc nécessaire de disposer de ressources compétentes et rapidement mobilisables pour réaliser les évolutions nécessaires
- Le système d'information étant entièrement basé sur un environnement Microsoft, il paraît cohérent d'utiliser une solution décisionnelle du même éditeur. De plus afin de tenir compte des ressources limitées de l'association le choix d'une solution intégrant l'ensemble des composants décisionnels est à privilégier à la mise en œuvre de plusieurs produits complémentaires
- Les investissements fait lors des tests de la solution à savoir l'acquisition des licences et la formation de membres du service informatique et du contrôle de gestion pourront être réutilisés dans le cadre de ce nouveau projet.

La proposition a été acceptée à la condition de valider la capacité d'EDENIS à réaliser ce projet en interne par la réalisation d'un proof of concept sur un périmètre limité.

## **III.2 Proof of concept**

Lors de la phase d'analyse une tâche manuelle, quotidienne et à faible valeur ajoutée pour la personne la réalisant a été identifiée comme critique. Il s'agit de la mise à disposition quotidienne du taux d'occupation des établissements à la direction.(Annexe 2)

La mise en place de ce rapport permettra de mettre en œuvre les étapes suivantes d'un projet de système décisionnel :

- Analyse de l'existant
- Analyse des sources de données
- Identification des règles de gestion
- Intégration de données
- Conception de tables dans une base de données
- Conception et paramétrage de rapports
- Tests et validations

L'objectif du POC n'est pas de valider les compétences des ressources sur la totalité des concepts d'un projet décisionnel mais de valider la capacité d'EDENIS à mener ce type de projet en interne.

### **III.2.1 Analyse de l'existant**

La mise à disposition quotidienne du taux d'occupation repose sur un export manuel des entrées/sorties par établissement effectué à partir du logiciel de gestion administrative du résident. Ce fichier est ensuite intégré dans une feuille de calcul mettant à jour les valeurs du jour par l'intermédiaire de macro.

L'onglet des résultats est ensuite imprimé en PDF pour être envoyé aux directeurs d'établissements.

### III.2.2 Source de données

La source de données est unique. Il s'agit de la base de données au format SQL du logiciel de gestion administrative du résident. La documentation de la base de données de ce progiciel ne nous a pas été communiquée par l'éditeur. La base de données possède 2317 tables, cependant l'analyse des requêtes intégrées au requêteur interne de l'outil a permis de facilement identifier les tables nécessaires à l'extraction des données ainsi que leurs relations.

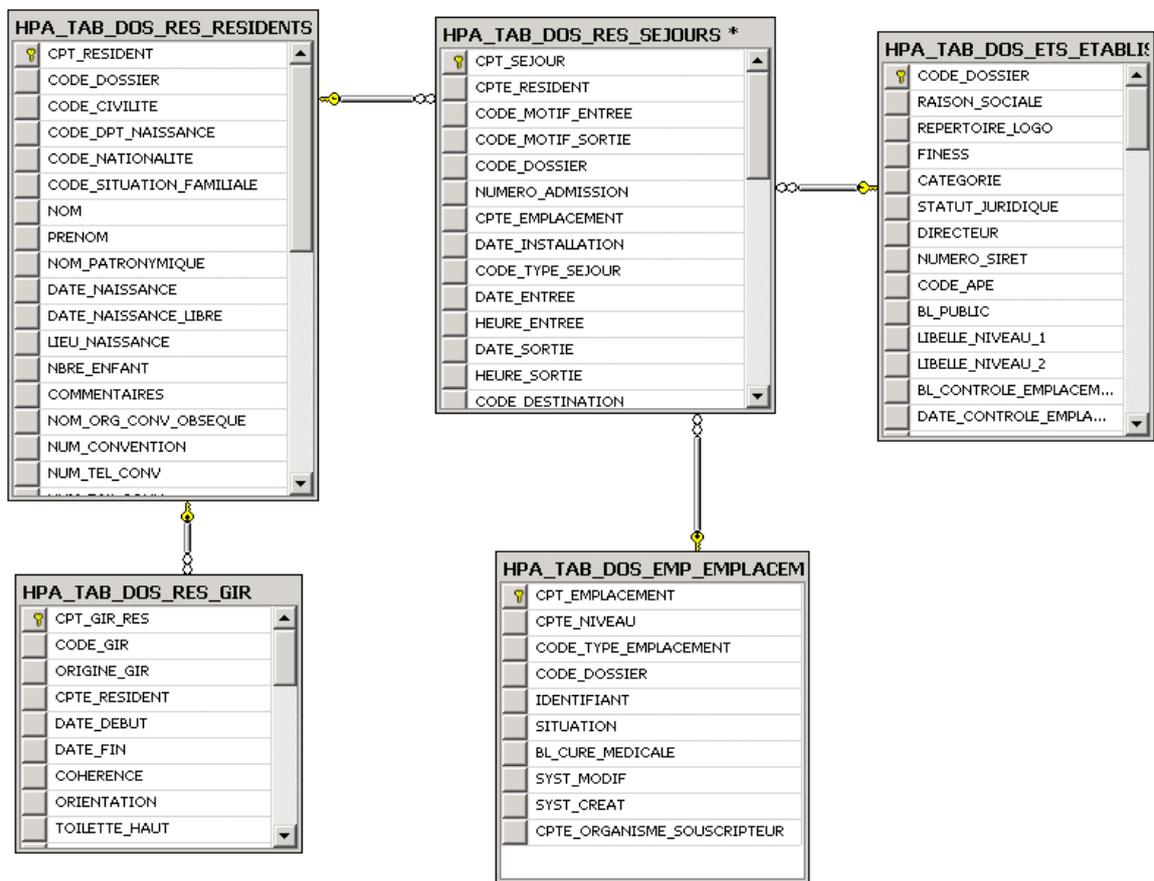


Figure 10: Schéma base de données gestion du résident

### III.2.3 Règles de gestion

L'interview du contrôleur de gestion lors de l'analyse de l'existant et les points réguliers lors de l'avancement de la conception ont permis de définir les règles de gestion suivantes : Les saisies des mouvements des résidents (entrées/sorties) ne se font pas forcément au jour le jour ce qui peut entraîner des écarts entre le rapport et la réalité du terrain. La gestion des dates est donc soumise aux règles suivantes

- Gestion des dates d'entrées :
  - Si la date du jour est antérieure à la date d'entrée saisie alors on retient la date d'entrée saisie sinon la date du jour
  - Si une date d'entrée saisie est postérieure à la date du jour et différente de la date d'entrée précédemment enregistrée, alors on met à jour la date d'entrée enregistrée.
- Gestion des dates de sorties :
  - Si une date de sortie est saisie postérieure à la date du jour alors on enregistre la date de sortie saisie.
  - Si une date de sortie est saisie antérieure ou égale à la date du jour alors on enregistre la date du jour.
  - Si une date de sortie postérieure à la date du jour est effacée alors on passe la date de sortie enregistrée à Null
  - Si une date de sortie saisie est postérieure à la date du jour et différente de la date de sortie précédemment enregistrée, alors on met à jour la date de sortie enregistrée.
- Gestion des annulations :
  - Si un séjour est annulé alors que sa date de début est antérieure à la date du jour alors la date de sortie est égale à la date du jour, sinon si la date de début est postérieure à la date du jour on définit date de début = date de fin.
- Gestion des IPP (Identifiant permanent du patient) :
  - A tous les nouveaux résidents est associé un numéro IPP, qui lui restera affecté pour tous les séjours dans tous les établissements du groupe qu'il sera amené à effectuer. Il faut s'assurer à chaque mise à jour de la cohérence des IPP.
- Règles de calculs

- Nombre d'entrées du jour : Nombre d'entrées égale à la date du jour
- Nombre de sorties du jour : Nombre de sorties égale à la date du jour
- Nombre de décès du jour : Nombre de dates de décès égale à la date du jour
- Nombre de présents du jour : Nombre de séjours pour lesquels la date d'entrée est inférieure ou égale à la date du jour et la date de sortie est nulle ou supérieure à la date du jour
- Capacité : Nombre de lits déclaré sur le site à la date du jour
- Capacité restante : capacité – Nombre de présents du jour
- Taux réalise N-4 N-3 N-2 N-1 : taux de remplissage réalisés des années précédentes.
- Taux objectif mensuel : taux objectifs du mois en cours
- Nombre de jours réalisés : Somme du nombre de présents du jour entre le premier jour du mois et la date du jour
- Nombre de jours restants : Nombre de jours entre la date du jour et le dernier jour du mois
- Nombre de jours manquants : Nombre de jours restants \* nombre de présents du jour
- Nombre de jour prévisionnels : Nombre de jours de réalisés + Nombre de jours manquants
- Nombre de Jour théorique : Capacité \* Nombre de jour du mois
- Taux d'occupation : Nombre de jours prévisionnels/Nombre de jours théoriques
- Nombre Entrées mensuelle : Somme des entrées entre le premier jour du mois et la date du jour
- Nombre de sorties mensuelle : Somme du nombre de sorties entre le premier jour du mois et la date du jour
- Nombre de Résidents soumis à TVA : Nombre de résidents dont la date d'entrée est postérieure au 01 janvier 2011 et dont la date de sortie est nulle ou postérieure à la date du jour
- Taux Résidents soumis à TVA : Nombre de résidents soumis à TVA divisé par le nombre de présents du jour
- Entrée previsionnel 15J : Nombre de résidents dont la date d'entrée est comprise entre la date du jour +1 et la date du jour +16

### III.2.4 Paramètres de mises à jour et de diffusion des rapports

Les données devront être intégrées quotidiennement après l'heure de départ des équipes administratives des EHPAD (20H). Les rapports devront être remis par mail le lendemain à 06 heures.

Les rapports seront diffusés au format PDF afin d'être non modifiables par les destinataires.

### III.2.5 Installation et développement.

Lors du lancement du proof of concept, la direction informatique n'avait pas mis en œuvre de serveur destiné au décisionnel. Une base de données nommée BDD\_EXPLOIT fut donc créée sur un serveur suffisamment dimensionné pour la recevoir.

Faute d'outil ETL disponible, l'alimentation de la base se fait au travers de procédures stockées exécutées automatiquement via une tâche de l'agent SQL serveur.

```
|  Procédures stockées système
|  dbo.01_AJOUT_NOUVEAUX_APPARTEMENT
|  dbo.02_AJOUT_NOUVEAUX_RESIDENTS
|  dbo.03_AJOUT_NOUVEAUX_SEJOURS
|  dbo.04_AJOUT_SORTIES
|  dbo.05_AJOUT_JOUR_TO
|  dbo.06_MAJ_DONNEE_A_UNE_DATE
|  " . . . . .
```

La base de données comprend une table de fait périodiques résumant quotidiennement et par établissement les informations suivantes :

- Nombre de présents
- Nombre d'entrées ; de sorties
- Nombre de sorties par motifs
- Objectif de remplissage de la période

### III.2.6 Tests

Durant tout le développement, chaque composant a été testé unitaire afin de valider les résultats obtenus vis-à-vis des règles de gestions.

Une fois que chaque composant a été validé, des tests d'intégration de l'ensemble des composants ont été menés afin de valider le fonctionnement global de la solution.

Enfin un test de validation en deux parties a été effectué :

Une validation fonctionnelle afin de valider les résultats obtenus, ce test a été réalisé par la vérification de la conformité des rapports générés par la solution à ceux du contrôle de gestion sur une période de plusieurs semaines.

Une validation technique afin de s'assurer de la stabilité, des performances de la solution ainsi que de la maîtrise de la volumétrie

A l'issue de ces tests une recette a été prononcée et la direction financière a validé la poursuite du projet avec pour objectif l'intégration de l'ensemble des sources de données de l'association.

### III.3 Gestion de projet

Le lancement du projet a été validé par la direction financière avec pour objectif d'intégrer à terme l'ensemble des sources de données de l'association.

La réalisation du projet devra être effectuée par les ressources internes en s'appuyant sur la suite de business intelligence de SQL serveur.

D'après les besoins et contraintes exprimés par le contrôle de gestion et la direction financière, il apparaît que le projet sera un projet majeur du système d'information de l'association et que celui-ci impactera fortement le service informatique et les utilisateurs métiers durant plusieurs mois. Face à ce constat de criticité, une gestion de projet rigoureuse est nécessaire.

Une organisation par phases entrecoupées de jalons est mise en place ainsi qu'un découpage par domaine permettant de réduire le volume et la complexité du projet en adéquation avec les ressources allouées.

Toutes les sources de données identifiées par le contrôle de gestion n'ayant pas la même urgence à être intégrées dans un entrepôt de données et afin de tenir compte des ressources humaines allouées à ce projet, il a été décidé de la mener avec une démarche incrémentale adaptée de la démarche conseillée par R. KIMBALL

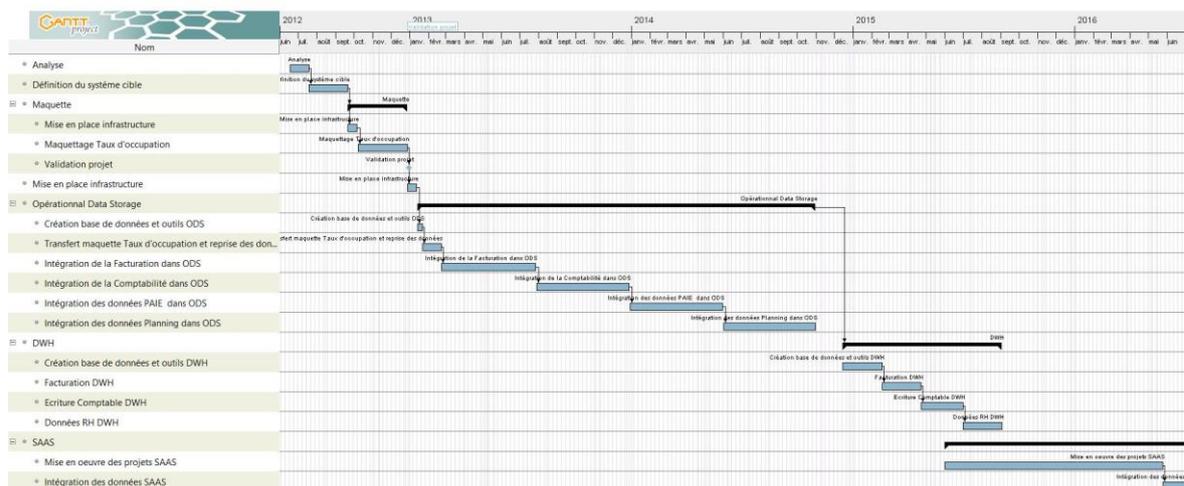


Figure 11 : Diagramme de Gantt Projet EDENIS

Des trois chemins parallèles du cycle de vie d'un projet décisionnel, seul deux seront traités à chaque itération (la modélisation et la conception des applications BI). Le chemin de l'architecture technique sera réduit à un contrôle des volumétries, la solution étant validée pour l'ensemble du projet.

Le projet sera ainsi découpé en plusieurs phases concernant des domaines de tailles réduites qui seront réalisées les unes après les autres selon la même méthodologie.

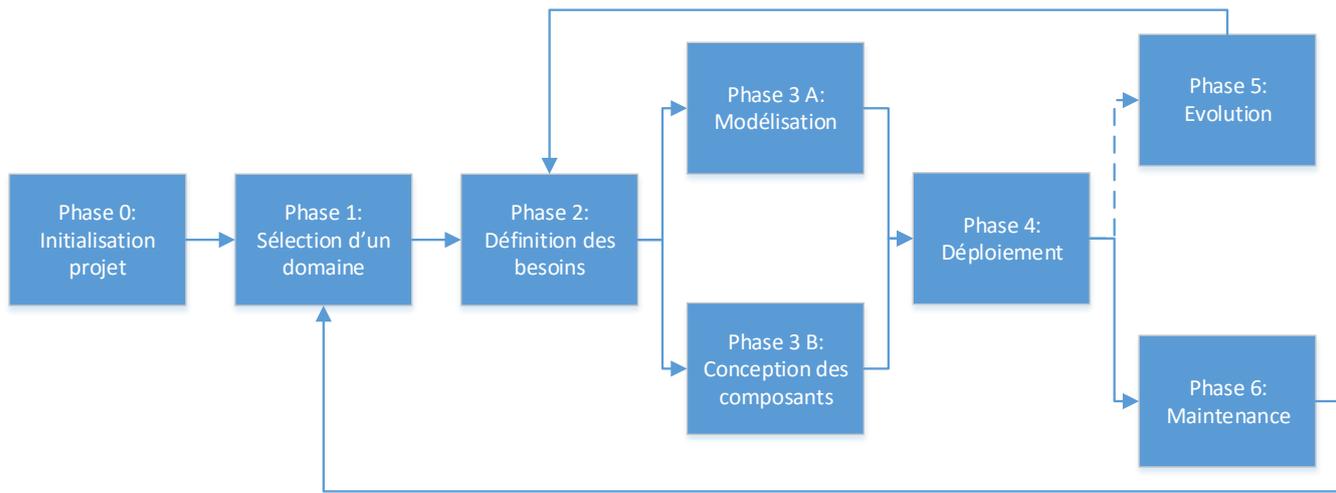


Figure 12 Diagramme phases projet

Cette démarche permettra d'éviter l'effet tunnel inhérent à tout projet important. Le contrôle de gestion bénéficiera de livraisons régulières et le choix du prochain domaine à traiter pourra être ajusté en fonction des ressources disponibles

## **III.4 Phase 0 : Initialisation projet**

### **III.4.1 Mise en place infrastructure et environnement de développement**

Deux environnements ont été mis en place un destiné à la production et un serveur de développement et de tests.

Le serveur décisionnel de production héberge :

- Une instance de serveur Microsoft SQL serveur : Cette instance pilote deux bases de données nommées ODS et DWH.
- Une installation du composant Microsoft SQL Server Intégration Services
- Une installation du composant Microsoft SQL Server Reporting Services
- Une installation du composant Microsoft SQL Server Analysis Services
- Une installation du composant Microsoft SQL Business Intelligence Management studio
- L'utilitaire d'exécution de package SSIS
- Un dossier partagé destiné à la collaboration projet. Celui est composé de quatre sous-dossiers :
  - Imports : dépôt des fichiers sources de données qui serviront à alimenter les ETL
  - Export : dépôt des fichiers exports notamment pour les exports programmés
  - Log : dépôt des fichiers LOG Project permettant le suivi et l'analyse de l'exécution des ETL.
  - Project : dépôt des sources de tous les éléments projets et notamment de tous les ETL.

### **III.5 Phase 1 : Sélection d'un domaine**

La sélection d'un domaine à traiter dans la solution décisionnelle se fait en concertation avec le contrôle de gestion et éventuellement la direction métier impliquée. Elle se base sur un arbitrage en fonction des points suivants :

- Criticité du besoin.
- Estimation du temps de travail.
- Disponibilité des ressources.

### **III.6 Phase 2 : définition des besoins**

Pour chaque domaine, il est nécessaire d'analyser les besoins utilisateurs afin de déterminer les solutions à mettre en œuvre. J'ai fait le choix d'utiliser la méthode de formulation des exigences gérées par les utilisateurs afin de cibler au mieux les besoins utilisateurs plutôt que d'analyser les documents existants car elle permet de concentrer les efforts vers les éléments réellement nécessaires plutôt que vers les éléments disponibles.

La formulation des exigences gérée par les utilisateurs est une méthode reposant sur la définition des exigences via l'examen des procédures exécutées par les utilisateurs au cours d'entretiens.

Lors de cette phase est également réalisé un inventaire des sources de données afin de vérifier leur disponibilité pour une intégration (lieu et format de stockage, mot de passe, droits d'accès, ...).

### **III.7 Phase 3 A : Modélisation des données.**

Pour chacune des sources de données identifiées, il est ensuite nécessaire de déterminer le modèle de données la composant. Pour les sources de données documentés, une relecture de la documentation permet de facilement identifier les données, leurs contraintes et leurs relations.

#### **III.7.1 Analyse des données existantes.**

Pour chaque domaine, il est nécessaire d'identifier les données produites par les opérateurs métiers et celles qui sont nécessaires à la production des rapports. Pour cela on s'appuiera sur la cartographie de couverture fonctionnelle de l'association qui permettra de déterminer si les données sont gérées dans une source de données maîtrisée du système d'information (source de données connue et documentée par le service informatique) ou dans une source non maîtrisée tel que des documents produits et gérés par les utilisateurs ou les bases de données pour lesquels les éditeurs ne fournissent pas de modèle de données.

Deux formats de sources de données fichier ont été identifiés au sein de l'association, les fichiers structurés de type csv et les fichiers tableurs.

Microsoft SQL serveur management studio propose une fonctionnalité d'import de données permettant à partir d'une source de données fichiers, d'analyser son contenu et de créer une table de destination contenant les colonnes aux formats identifiés. La cohérence des formats des colonnes de la table de destination est dépendante de la déclaration des formats dans le fichier source. Un ajustement manuel est donc souvent nécessaire notamment sur les longueurs des champs.

Pour les sources de données de type base de données SQL il est inenvisageable de procéder à une copie totale de la base. En effet une base de données de progiciel est composée de nombreuses tables inutiles dans le cadre d'un projet décisionnel (tables système, tables temporaires, tables de fonctionnalité non mise en service, ...). De plus avoir accès au contenu d'une base de données ne signifie pas qu'il est possible d'exploiter ces données dans un projet décisionnel.

Les bases de données ont donc toutes été documentées avant leurs intégrations dans l'ODS, pour certaines cela ne pose pas de difficulté, l'éditeur fournissant la documentation nécessaire avec les livrables ou sur demande, mais dans la majorité des cas ces informations n'ont pas pu m'être transmises.

Le modèle physique de la base peut être retrouvé par l'analyse des clefs permettant d'identifier les relations entre les tables. Cela peut se faire au moyen de plusieurs outils comme les fonctionnalités de diagramme de SQL serveur qui affiche l'ensemble des tables jointes à une table ciblée

Connaitre la structure de la base ne permet pas d'exploiter les données qui y sont stockées. Le nombre de tables est fréquemment de l'ordre de la centaine et celui des champs du milliers avec une nomenclature propre à l'éditeur. Pour déterminer le modèle conceptuel de la base, il donc nécessaire de réaliser une mise en correspondance des données attendues par EDENIS avec celles stockées dans la base.

Plusieurs méthodes de retro ingénierie des bases ont été employées selon les cas :

- La base de données est de faible dimension et le nommage des tables et des champs est explicite :

Une exploration manuelle de la base est faite afin de déterminer pour chaque donnée la table et le champ associé. On documente pour chaque donnée le nom de la table, du champ et le format de stockage.

- La base de données est associée à une logiciel de Reporting permettant de visualiser le code des requêtes effectuées :

Un relevé des tables et des champs dans les tables permet de faire la mise en correspondances avec les données restituées par les requêtes et ainsi documenter la base

- La base de données utilise une convention de nommage non explicite et aucune information n'est disponible au travers du logiciel de Reporting ou celui-ci n'existe pas :

Dans l'application métier, on affiche une donnée recherchée puis on affiche les logs d'exécution de la requête par le serveur SQL via un script (Annexe 6)

### **III.7.2 Modélisation ODS**

La difficulté de la création de l'ODS réside dans la nécessité de maîtriser les sources de données. Une fois la source de données documentée, il est possible de concevoir les tables nécessaires à l'ODS selon la méthode suivante.

A partir du dictionnaire des données, on sélectionne toutes les tables contenant les données attendues

A partir du modèle physique on ne retient que les clefs reliant les tables sélectionnées. Dans le cas où des tables sélectionnées ne possèdent aucune relation directe on analyse le modèle physique des données afin d'ajouter les tables nécessaires aux jointures.

Afin de ne pas traiter des données inutiles, on élimine les colonnes non nécessaires.

On dispose à présent du modèle physique des tables à ajouter à l'ODS, le code de génération est rédigé puis exécuté sur le serveur décisionnel

### **III.7.3 Conception des ETL ODS**

L'ODS étant uniquement destinée à centraliser les données métiers sans mettre en œuvre de mécanismes d'historisation, de contrôle des données, ou de transformation de schéma ; les ETL se composent typiquement de deux tâches. Tout d'abord l'effacement de toutes les données contenues dans les tables puis leurs rechargements complets.

L'exécution des ETL d'alimentation est automatisée via une tâche d'agent SQL serveur selon la planification d'actualisation demandée.

### III.7.4 Modélisation DATA WAREHOUSE

La démarche suivante a été mise en œuvre :

Avant le lancement de la construction du Data warehouse, j'ai commencé par établir une cartographie des processus métier de l'entreprise.

A partir de ce document, j'ai identifié les axes d'analyses communs de l'association afin de déterminer les dimensions communes à plusieurs des domaines qui seront intégrés par la suite.

Ces dimensions génériques ont été réalisées puis soumises à la validation du contrôle de gestion.

- Dimension Temps (Annexe 5)
- Dimension Etablissement
- Dimension Résident
- Dimension Compte comptable
- Dimension salariée

Pour chaque domaine traité, on identifie des faits nécessaires à partir de l'analyse des besoins. Pour chacun de ces faits est identifié l'ensemble des dimensions associées nécessaire à l'analyse.

Lors de cette phase est également déterminé le niveau des informations détaillées dont le modèle dimensionnel peut disposer : la granularité de chaque table

Une fois les faits, les dimensions et la granularité identifiés, le modèle physique de chaque étoile est réalisé.

### III.7.5 Conception ETL DWH

Les ETL destinés au Data warehouse sont plus complexes que ceux de l'ODS car ils impliquent des modifications des schémas de données.

Le processus suivant a été défini et appliqué pour l'alimentation de chacune des tables de fait :

1. Chargement et mises à jour de toutes les tables de dimension reliées à la table de fait
  - a. Extraction des données de l'ODS
  - b. Conversion et nettoyage des données
  - c. Recherche des non correspondances
  - d. Génération des méta données
  - e. Ajout des nouveaux enregistrements
  - f. (Optionnel) Mise en œuvre de mécanisme de gestion des dimensions à variations lentes
  - g. Gestion des erreurs et audits
2. Alimentation de la table de fait
  - a. Extraction des données de l'ODS
  - b. Conversion et nettoyage des données
  - c. Recherche des non correspondances
  - d. Substitution des clefs
  - e. Génération des méta données
  - f. (Optionnel) Insertion des champs calculés
  - g. (Optionnel) Agrégation des données
  - h. Ajout des nouveaux enregistrements
  - i. Gestion des erreurs et audits

### **III.8 Phase 3 B : Conception des composants de la solution décisionnelle afin de répondre aux problématiques métiers.**

Une solution décisionnelle met à disposition des responsables de projets de nombreux composants et méthodes destinés à répondre aux problématiques métiers. Avant de déployer une solution décisionnelle, il est nécessaire de définir quelles problématiques vont être résolues par quels composants et selon quelles méthodes, afin de créer une politique du système d'information décisionnel adaptée à l'association.

J'ai donc conçu une méthodologie de mise en œuvre des composants décisionnels pour chacun des objectifs et des besoins exprimés du projet.

#### **III.8.1 Problématique de fragmentation du système d'information**

##### **III.8.1.1 Objectifs**

La principale problématique engendrée par la fragmentation du système d'information est la difficulté d'accès aux données dans un objectif de Reporting. La dépendance aux fonctionnalités des applications métiers est totale, leurs possibilités sont très variables voire dans certains cas inexistantes.

L'objectif de ce composant est de fournir des capacités de Reporting au travers d'un outil commun pour l'ensemble des données de l'association

##### **III.8.1.2 Mise en œuvre**

Cette problématique a été résolue par la mise en œuvre d'une base de données de type « Operational Data Store » (ODS).

Le principe de l'ODS est d'intégrer dans une base de données unique ; les données de l'ensemble des sources de production.

L'alimentation de la base de données ODS consiste en la création des tables et des ETL modélisées selon la méthodologie vue au point III.7.2 et III.7.3 .

L'ODS devient ainsi la source de données unique de l'entreprise et permet l'utilisation des mêmes outils de Reporting pour l'ensemble des données répondant ainsi aux difficultés de maitriser des multiples outils de Reporting existant au sein du système d'information.

L'ODS a aussi vocation à être la source de données unique du Data warehouse d'EDENIS.

### **III.8.1.3 Résultats**

La mise en œuvre de l'ODS a permis la création de rapports de contrôles des données saisies dans les logiciels métiers permettant la mise en œuvre de contrôle complémentaires et indépendants non disponibles dans les applications. L'ODS est cependant limité par la dépendance de sa structure de données à celles des applications métiers, il ne permet pas de croiser des données provenant d'applications différentes ni de mettre en œuvre un mécanisme d'historisation.

Exemple :

L'ODS a permis de mettre en place un contrôle des dates de modification des tarifs de certaines prestations fixées réglementairement par publication au journal officiel. Ce type de contrôle a permis d'être proactif sur les erreurs de facturation et de réduire le nombre d'avis à réaliser. Il ne permet cependant pas de mettre en place d'alertes de suppression de tarifs

## **III.8.2 Problématique d'indépendance au système d'information et historisation des données**

### **III.8.2.1 Objectifs**

L'objectif est de mettre en œuvre un composant permettant le stockage des données de manière indépendante du système d'information actuel et de ses évolutions.

### **III.8.2.2 Mise en œuvre**

La solution mise en œuvre consiste à intégrer les données dans un Data warehouse. Les données sont intégrées selon un schéma en étoile, dans un entrepôt de données multi dimensionnel en utilisant des dimensions conformes entre des tables de fait permettant d'effectuer des croisements avec les autres étoiles de la base.

L'alimentation de la base de données DWH consiste en la création des tables et des ETL modélisées selon la méthodologie vue aux points III.7.4 et III.7.5 .

Selon les besoins exprimés les faits seront stockés sous forme détaillés ou agrégés et l'historisation des données sera mise en place.

Cette phase permet la mise en place de rapports dynamiques avec des fonctionnalités avancées d'explorations des données.

### **III.8.2.3 Résultats**

La mise en œuvre du Data warehouse a permis d'affranchir l'analyse des données de toutes dépendances au système d'information. Edenis possède désormais deux systèmes d'information, à savoir le système d'information de gestion SIG et le système d'information décisionnel SID qu'il convient de faire évoluer de manière conjointe.

### **III.8.3 Problématique de consolidation des données**

#### **III.8.3.1 Objectifs**

La fragmentation du système d'information d'EDENIS engendre une impossibilité de croiser les données provenant des sources différentes. Il était donc nécessaire de mettre en œuvre des mécanismes de réconciliations des nomenclatures de données.

#### **III.8.3.2 Mise en œuvre**

Plusieurs méthodes ont été conçues selon le composant ou la consolidation des données qui devait être mis en œuvre.

- Au niveau de l'ODS

Au niveau de l'ODS, pour certains rapports, des tables de correspondances ont été mises en œuvre afin de permettre les jointures sur des données provenant de sources de données différentes.

- Au niveau du Data warehouse

La consolidation des données, est un des objectifs d'un Data warehouse, qui est réalisé par l'utilisation des dimensions partagées et la création d'étoiles thématiques contenant des données de sources différentes.

- Au niveau des rapports

Des jeux de données incluant des jointures, par l'intermédiaire des champs calculés ont également été mis à disposition, notamment afin de permettre l'interrogation de tables de l'ODS contenant des informations de dates, stockées dans des formats différents.

#### **III.8.3.3 Résultats**

La mise en œuvre de ces composants permet aux utilisateurs la réalisation de croisements de données à tous les niveaux du système décisionnel, et ceci sans être dépendant de la fin de la mise en œuvre du Data warehouse, évitant ainsi un effet tunnel, ce composant étant le plus complexe à définir.

### III.8.4 Comment résoudre la problématique de mise à disposition des données aux utilisateurs ?

#### III.8.4.1 Objectifs

Le stockage des données dans l'ODS ou le Data warehouse permet de disposer de toutes les données de l'entreprise dans un entrepôt centralisé. Cependant seuls des utilisateurs ayant le profil informaticien peuvent effectuer des requêtes et des traitements sur ces données. Il convient donc de mettre à disposition des utilisateurs métiers des outils d'accès aux données plus adaptés.

#### III.8.4.2 Mise en œuvre

Pour cela on s'appuiera sur la suite SSRS fournissant des fonctionnalités de portail d'entreprise et de création de rapports d'entreprise qui permettant de concevoir, créer, déployer et gérer des rapports,

#### III.8.4.3 Mise en place du portail décisionnel

Le portail décisionnel repose sur un serveur IIS7 installé sur le serveur décisionnel

The screenshot shows the configuration interface for the Web Service URL in SSRS. It includes a title bar 'URL du service Web', a globe icon, and a descriptive text: 'Configurez une URL utilisée pour accéder au serveur de rapports. Cliquez sur Options avancées pour définir plusieurs URL pour une seule instance du serveur de rapports ou pour spécifier d'autres paramètres sur l'URL.' Below this are three main sections: 1. 'Répertoire virtuel du service Web Report Server' with a text box containing 'ReportServer\_DECIS\_PRD'. 2. 'Identification du site du service Web Report Server' with fields for 'Adresse IP' (dropdown: 'Assigné (recommandé)'), 'Port TCP' (text box: '80'), 'Certificat SSL' (dropdown: '(Non sélectionné)'), and 'Port SSL' (text box). An 'Options avancées...' button is located to the right. 3. 'URL du service Web Report Server' with a text box containing the URL 'http://SRV-DECIS:80/ReportServer\_DECIS\_PRD'.

Le portail est mis à disposition de tous les utilisateurs via la diffusion d'un raccourci sur leur bureau par une GPO.

**Raccourci Décisionnel**  
Données recueillies le : 22/04/2016 14:23:33 [afficher tout](#)

**Configuration ordinateur (désactivée)** [masquer](#)

Aucun paramètre n'est défini.

**Configuration utilisateur (activée)** [masquer](#)

**Préférences** [masquer](#)

**Paramètres Windows** [masquer](#)

**Raccourcis** [masquer](#)

**Raccourci (chemin d'accès : %DesktopDir%\Portail Décisionnel)** [masquer](#)

**Portail Décisionnel (ordre : 1)** [masquer](#)

**Général** [masquer](#)

Action	Mettre à jour
<b>Attributs</b>	
Type de cible	URL
Chemin de raccourci	%DesktopDir%\Portail Décisionnel
URL cible	http://srv-decis/Reports_DECIS_PRD/Pages/Folder.aspx
Chemin d'accès à l'icône	%SystemRoot%\system32\SHELL32.dll
Index de l'icône	288
Touche de raccourci	None
Exécuter	Fenêtre normale

**Commun** [masquer](#)

**Options**

Interrompre le traitement des éléments sur cette extension si une erreur se produit sur cet élément	Non
Exécuter dans le contexte de sécurité de l'utilisateur connecté (option de la stratégie utilisateur)	Oui
Supprimer cet élément lorsqu'il n'est plus appliqué	Non
Appliquer une fois et ne pas réappliquer	Non

Figure 13GPO diffusion raccourci portail décisionnel

### III.8.4.4 Création des rapports

Les rapports sont développés via Visual Studio Business Intelligence pour être publiés sur le portail décisionnel.

Un rapport est une mise en forme des données d'un dataset en fonction de paramètres et de variables.

### III.8.4.5 Résultats

Les utilisateurs ont désormais à leur disposition un portail décisionnel à partir duquel ils peuvent exécuter les rapports disponibles.

## III.8.5 Comment répondre au besoin d'autonomie pour la création de rapports ?

### III.8.5.1 Objectifs

Les utilisateurs avancés doivent pouvoir créer ou modifier et publier des rapports de manière autonome.

### III.8.5.2 Mise en œuvre

Pour cela on s'appuiera sur la suite SSRS fournissant un outil de création de rapports d'entreprise permettant de concevoir, créer, déployer et gérer des rapports, à savoir report builder 3.0. Celui-ci est mis à disposition au travers du portail décisionnel qui selon le profil des utilisateurs autorise l'installation de ce composant et la création de rapports.

### III.8.5.3 Accès aux données

Avant toutes créations de rapports il est nécessaire de fournir les accès aux données. Cela se fait au moyen d'un élément de type source de données qui est un objet définissant le chemin d'accès à une source de données et les informations d'authentications associées.

SQL Server Reporting Services  
**Nouvelle source de données**

Nom : DATAWAREHOUSE

Description : Source de données DATAWAREHOUSE  
production EDENIS

Masquer en mode Mosaïque

Activer cette source de données

Type de source de données : Microsoft SQL Server

Chaîne de connexion : Data Source=srv-decisisdecis\_prd;Initial  
Catalog=DWH

Se connecter avec :

Informations d'identification fournies par l'utilisateur qui exécute le rapport

Afficher le texte suivant pour demander à l'utilisateur un nom d'utilisateur et un mot de passe :

Tapez ou entrez un nom d'utilisateur et un mot de passe pour accéder

Utiliser comme informations d'identification Windows lors de la connexion à la source de données

Informations d'identification stockées en sécurité dans le serveur de rapports

Nom d'utilisateur : [ ]

Mot de passe : [ ]

Utiliser comme informations d'identification Windows lors de la connexion à la source de données

Emprunter l'identité de l'utilisateur authentifié une fois la connexion établie à la source de données

Sécurité intégrée de Windows

Figure 14Création d'une source de données

A partir d'une source de données il est possible de déclarer une vue de données au travers d'un élément de type dataset qui représente le résultat de l'exécution d'une requête sur une source de données.

Un concepteur graphique est disponible dans reports builder cependant pour les requêtes contenant des jointures complexes, l'utilisation de requêtes SQL est privilégiée.

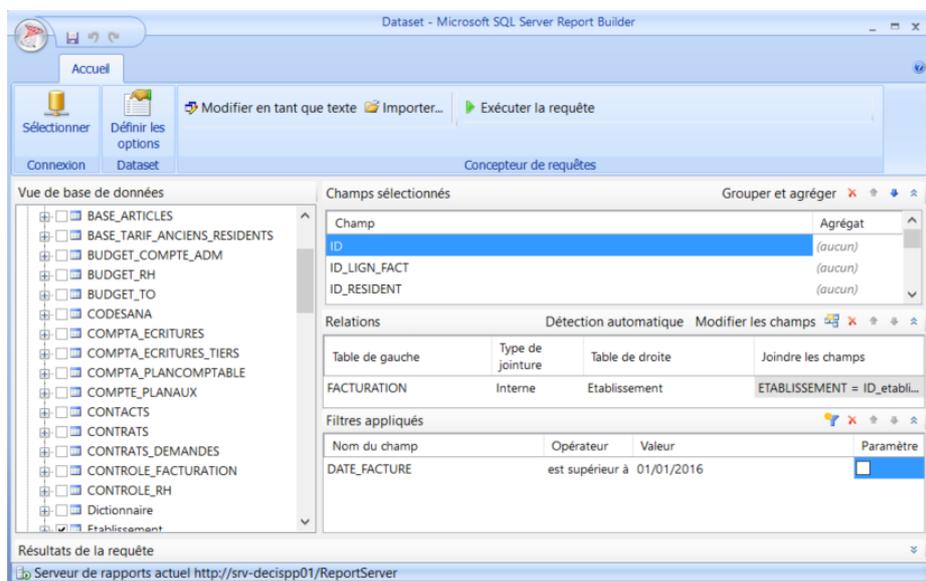


Figure 15 Création de dataset

Les sources de données et les dataset sont partagés au niveau du portail dans deux dossiers spécifiques afin d'être réutilisables par l'ensemble des rapports et des utilisateurs.

Les sources de données et les dataset peuvent également être intégrés directement dans certains rapports.

### III.8.5.4 Création des rapports

La création d'un rapport se déroule selon les étapes suivantes :

- Sélection d'une source de données et d'un dataset

Report Builder permet la création directe de source de données et de dataset ou la sélection d'éléments existants au niveau du portail décisionnel

- Traitement des données

Réalisation d'opérations sur les données remontées par le dataset afin de les préparer à l'affichage

- Filtre
- Paramètre
- Champs calculés
- Agrégation
- Fonctions

- Mise en forme des données
  - Table : liste le contenu qui permet de regrouper les lignes sous forme de sous totaux
  - Tableau matriciel : permet d'afficher le détail des regroupements
  - Liste : Affichage détaillé des données
  - Graphe : affichage sous forme visuelle des données

### **III.8.5.5 Résultats**

Les utilisateurs avancés de l'association ont désormais à leur disposition un outil de Reporting convivial, utilisable par des non informaticien. Il leurs permet d'élaborer des rapports à partir des sources de données et des dataset existants et de personnaliser des rapports ou d'en créer des nouveaux pour les déployer sur le portail décisionnel.

## III.8.6 Comment répondre à la problématique de diffusion des rapports par mail ?

### III.8.6.1 Objectifs

Tous les utilisateurs n'ont pas l'autorisation ou ne peuvent pas utiliser le portail web, l'envoi de rapports par mail permettrait néanmoins de leur diffuser de l'information.

### III.8.6.2 Mise en œuvre

La fonctionnalité d'envoi de mails existante dans SSRS a donc été mise en œuvre.

Le premier point a été de s'assurer de l'accessibilité du serveur SMTP par le serveur décisionnel, au moyen des commandes Telnet suivantes

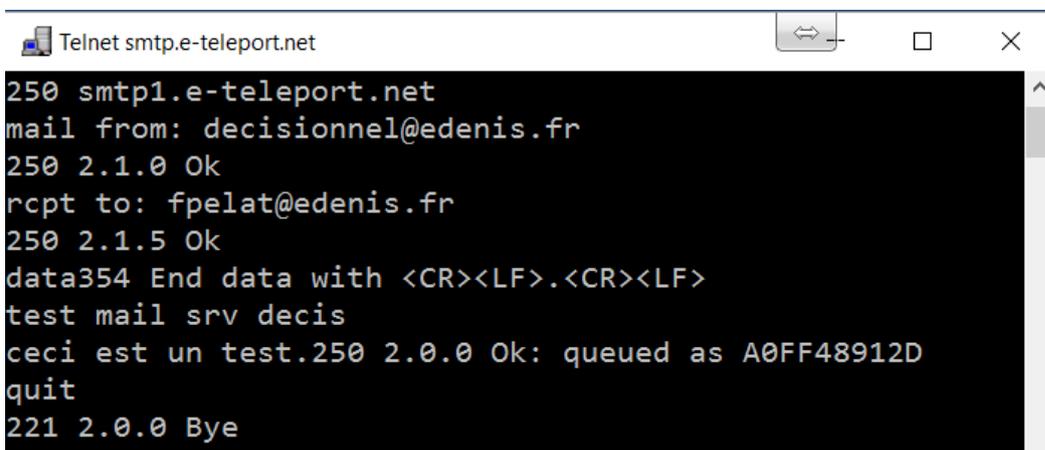
A screenshot of a Telnet window titled 'Telnet smtp.e-teleport.net'. The window shows a text-based SMTP session. The user enters 'mail from: decisionnel@edenis.fr', 'rcpt to: fpelat@edenis.fr', and 'data354 End data with <CR><LF>.<CR><LF>'. The server responds with '250 smtp1.e-teleport.net', '250 2.1.0 Ok', '250 2.1.5 Ok', and '250 2.0.0 Ok: queued as A0FF48912D'. The user enters 'test mail srv decis' and 'quit', and the server responds with '221 2.0.0 Bye'.

Figure 16: Commandes TELNET test serveur SMTP

La communication SMTP étant validée, les paramètres suivants de messagerie ont pu être renseignés au niveau du gestionnaire de configuration de Reporting services

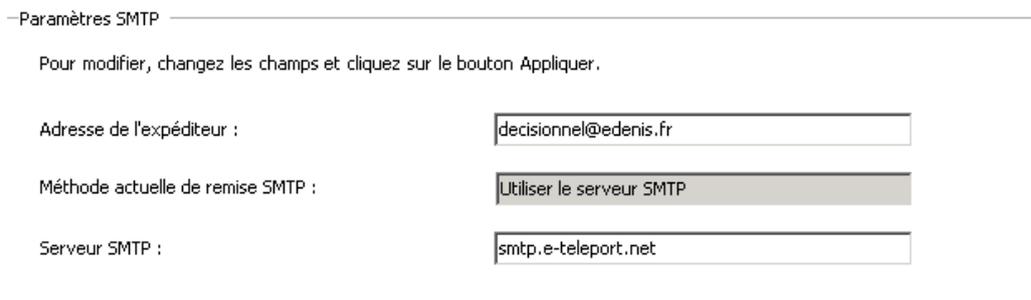
A screenshot of the 'Paramètres SMTP' configuration page in SSRS. The page has a title bar 'Paramètres SMTP' and a subtitle 'Pour modifier, changez les champs et cliquez sur le bouton Appliquer.' Below this, there are three configuration fields: 'Adresse de l'expéditeur :' with the value 'decisionnel@edenis.fr', 'Méthode actuelle de remise SMTP :' with the selected option 'Utiliser le serveur SMTP', and 'Serveur SMTP :' with the value 'smtp.e-teleport.net'.

Figure 17 Paramètre SMTP SSRS

Pour chaque rapport existant il est possible de définir un abonnement selon une planification ainsi que les modalités d'accès au rapport (pièce jointe, format, lien, ...)

**Options de remise de rapport**  
 Spécifiez les options de remise de rapports.

Remis par :

À :

Cc :

Cci :

Répondre à :

Objet :

Inclure un rapport    Format du rendu :

Inclure un lien

Priorité :

Commentaire :

**Options de traitement d'abonnement**  
 Spécifiez les options de traitement des abonnements.

Exécuter l'abonnement :

Lorsque l'exécution planifiée du rapport est terminée.   
 À 08:00, chaque lun de chaque semaine, à partir du 18/04/2016

Suivant une planification partagée :

À 19:20, le(s) 7 de chaque mois, à partir du 23/04/2015

Figure 18 Options de remise de rapport SSRS

Pour chaque abonnement il est également possible de définir les valeurs des paramètres à appliquer pour chacun des rapports afin de cibler au mieux leurs contenus et le profil des destinataires.

**Valeurs de paramètre de rapport**

Spécifiez les valeurs de paramètre de rapport à utiliser avec cet abonnement.

**mois du rapport**

Utiliser la valeur par défaut

**Annee du rapport**

Utiliser la valeur par défaut

**Etablissement du rapport**

Utiliser la valeur par défaut

Figure 19 Option de paramètres de rapport SSRS

**III.8.6.3 Résultats**

La mise en place de la fonctionnalité d'envoi de rapport par mail a permis aux utilisateurs avancés d'automatiser et de maîtriser la diffusion d'informations en interne et externe.

### **III.8.7 Comment répondre à la problématique d'intégration de données non structurées ?**

#### **III.8.7.1 Objectifs**

Edenis utilise plusieurs trames de fichiers non structurées tel que des fichiers Word ou bien Excel mais utilisant des contrôles de formulaires. Leurs intégrations directes par un ETL n'est pas possible et une solution de contournement n'impactant pas les utilisateurs doit être mise en place.

#### **III.8.7.2 Mise en œuvre**

Ces trames utilisent des scripts VBA (Annexe 7) pour la mise à jour des champs et autres fonctions internes. Des scripts complémentaires ont été développés afin qu'au moment de la sauvegarde des documents les données soient exportées dans une base de données structurées (xls) créée pour cet usage.

### **III.8.8 Résultats**

La mise en œuvre de ces scripts et de ces bases de données complémentaires a par exemple permis d'intégrer les données des formulaires de demandes de contrat (environ 300 demandes mensuelles) et de mettre en œuvre des rapports de contrôles et des statistiques sans impacter les habitudes des utilisateurs de ces trames.

## III.8.9 Comment déclencher manuellement un package ETL ?

### III.8.9.1 Objectifs

Les utilisateurs avancés tel que le contrôleur de gestion doivent pouvoir déclencher manuellement l'exécution des ETL

### III.8.9.2 Mise en œuvre

J'ai déployé sur les postes de ces utilisateurs l'utilitaire d'exécution de package qui permet de déclencher l'exécution des packages SSIS stocké sur le système de fichier du serveur décisionnel.

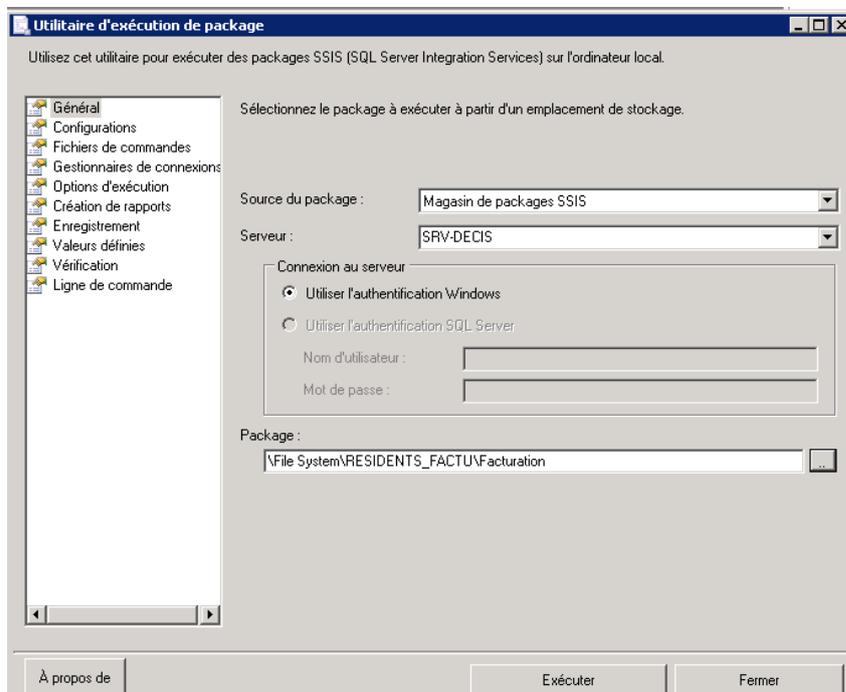


Figure 20 Utilitaire d'exécution de package

### III.8.9.3 Résultats

L'utilisation de ce composant permet au contrôleur de gestion une totale autonomie dans la mise à jour des données et notamment l'exécution d'un import correctif suite à la détection d'une erreur de saisie. C'est par exemple le cas d'une erreur de facturation qui sera détectée par un rapport de contrôle basé sur l'ODS, l'opérateur réalisera une annulation et une nouvelle génération de facture qui pourra être prise en compte par un nouvel import sans attendre l'exécution de la planification suivante.

### **III.9 Phase 3 B : Conception des composants de la solution décisionnelle afin de répondre aux problématiques d'exploitation.**

De la même manière que la mise en œuvre d'une solution décisionnelle nécessite de définir une méthodologie de mise en œuvre pour chacun des objectifs métier, il est également nécessaire de définir des méthodologies de conception des composants afin de répondre aux problématiques d'exploitation informatique de la solution.

#### **III.9.1 Gestion des erreurs ETL ?**

Malgré le soin apporté à la conception des ETL, des erreurs peuvent aboutir à un chargement partiel des données voire à la non-exécution de l'ETL. Ces erreurs doivent donc être détectées et tracées afin de permettre l'identification et la correction de l'origine et le rétablissement éventuel des données.

##### **III.9.1.1 Alertes d'exécution des ETL**

Un mécanisme d'alerte de l'administrateur a été mis en place sous la forme de tâche d'envoi de mail, en cas de détection d'erreur d'exécution d'une tâche d'un ETL ou de l'ensemble de l'ETL.

##### **III.9.1.2 Log des erreurs de traitement de données.**

Afin de faciliter la correction des erreurs détectées toutes les lignes n'ayant pu être traitées de bout en bout par un ETL sont redirigées dans un fichier de LOG.

Cela à deux avantages :

- Permettre d'identifier rapidement les sources d'erreurs car les données et les codes erreurs sont enregistrées ensemble.
- Pouvoir communiquer aux utilisateurs l'impact de l'erreur sur les données de leurs rapport (une erreur peut ne pas avoir d'impact sur un rapport car les lignes ne concernent pas les mêmes périodes)

## **III.9.2 Comment maîtriser la volumétrie et les performances ?**

### **III.9.2.1 Objectifs**

Les utilisateurs ont exprimé des exigences en termes de fréquence de rafraîchissement de données et de délai de mise à disposition. Ces délais doivent être respectés malgré l'augmentation des volumétries et du périmètre fonctionnel sous peine d'un abandon de l'utilisation de la solution par les utilisateurs

### **III.9.2.2 Mise en œuvre**

La solution décisionnelle gère et restitue de grandes quantités de données, il est nécessaire de mettre en place des mécanismes et des bonnes pratiques sur chacun des composants de la solution décisionnelle afin de garantir des temps de traitements et de restitutions acceptables par les utilisateurs.

L'optimisation se fait notamment au niveau des bases de données, des ETL et des rapports.

#### **III.9.2.2.1 Optimisation des bases de données**

La mise en œuvre d'index sur les clefs étrangère des tables des bases de données ODS et DWH permet d'accélérer le traitement des jointures des requêtes.

Les index sont également déclarés sur les clefs métiers afin d'accélérer les opérations de recherches et les jointures des ETL.

Un plan de maintenance est également mis en œuvre sur ces bases de données, il réalise les tâches suivantes :

- Tâche réduire la base de données
- Tâche vérifier l'intégrité de la base de données
- Optimisation de l'index :
  - Si le taux de fragmentation est  $>5\%$  et  $<30\%$  : Tâche réorganiser l'index
  - Si le taux de fragmentation est  $\geq 30\%$  : Tâche reconstruire l'index
- Tâche de nettoyage d'historique
- Tâche de nettoyage de maintenance

### III.9.2.2.2 Optimisation des ETL

Les ETL sont des outils très souples permettant d'effectuer de nombreuses tâches, cependant leurs performances sont très dépendantes de leurs conceptions. Les bonnes pratiques suivantes ont été mises en œuvre afin d'optimiser l'exécution des ETL :(Kumar, 2015)

- Il est très pénalisant d'importer la totalité des sources de données. Restreindre la sélection des données à importer aux champs nécessaires aux traitements permet d'économiser les ressources systèmes.
- Lorsque les sources de données sont de types fichiers, SSIS met en œuvre une reconnaissance de type automatique, ce typage entraîne une surconsommation de la mémoire car la plupart des colonnes seront typées comme des chaînes de 255 caractères. Les types des champs doivent donc être déclarés explicitement.
- Lorsque plusieurs sources de données non dépendantes sont utilisées, on met en œuvre l'extraction en parallèle des données dans le flux de contrôle.
- Le choix des types de composants de transformation est primordial. Une transformation à sorties synchrones traite chacune des lignes d'entrée lorsqu'elles traversent le composant. Une transformation à sorties asynchrones attend d'avoir reçu toutes les lignes d'entrée avant de procéder au traitement des données. Les traitements asynchrones sont très pénalisants en termes de temps de traitements et à n'utiliser qu'en cas de nécessité comme la réalisation de tri et d'agrégation de données.
- Pour les chargements de grand volume de données il faut veiller aux caractéristiques de destination. Une table indexée ralentira fortement les opérations d'insertion, de mise à jour et de suppression. L'ajout de tâches de suppression et de recréation d'index sera envisagé au cas par cas.
- La sélection des options du composant de destination OLEDB influence fortement les performances, il convient de privilégier l'option chargement rapide des données et d'y associer les options suivantes : verrouillage de la table et ne pas vérifier les contraintes.

### III.9.2.2.3 Optimisation de l'affichage des rapports.

La génération des rapports est une tâche consommatrice de ressources système, l'exécution parallèle de nombreux rapports ou l'exécution de certains rapports complexes peut nécessiter des temps de traitements de plusieurs minutes.

Selon les besoins utilisateurs deux options ont été mises en œuvre, la mise en cache du rapport ou la création d'un instantané.

La mise en cache d'un rapport consiste, après son affichage, à stocker en mémoire une copie temporaire du rapport et à lui définir une durée d'expiration.

Toujours exécuter ce rapport avec les données les plus récentes

Ne pas mettre en cache les copies temporaires de ce rapport

Mettre en cache une copie temporaire du rapport. Faire expirer la copie du rapport après un certain nombre de minutes :

Mettre en cache une copie temporaire du rapport. Faire expirer la copie du rapport selon la planification suivante :

Planification spécifique aux rapports   
À 08:00, chaque lun de chaque semaine, à partir du 27/04/2016

Planification partagée   
À 06:00, chaque mar de chaque semaine, à partir du 22/12/2014 et jusqu'au 14/12/2015

Figure 21 Option de mise en cache d'un rapport

La création d'un instantané de rapport consiste à planifier un pré calcul d'un rapport qui sera ainsi mis à disposition des utilisateurs jusqu'à la prochaine planification.

Effectuer le rendu de ce rapport à partir d'un instantané de rapport

Utiliser la planification suivante pour créer des instantanés de rapport :

Planification spécifique aux rapports   
À 08:00, chaque lun de chaque semaine, à partir du 27/04/2016

Planification partagée   
À 01:00, le(s) 21 de chaque mois, à partir du 23/04/2015

Créer un instantané du rapport lorsque vous cliquez sur le bouton Appliquer de cette page

Figure 22 Option de création d'instantanés de rapport

Ces deux mécanismes peuvent accélérer fortement le temps d'affichage des rapports tout en étant transparents du point de vue des données qu'ils contiennent, pour cela il faut calquer leurs planifications à celle de mise à jour des données par les ETL.

### III.9.3 Comment gérer les autorisations ?

#### III.9.3.1 Objectifs

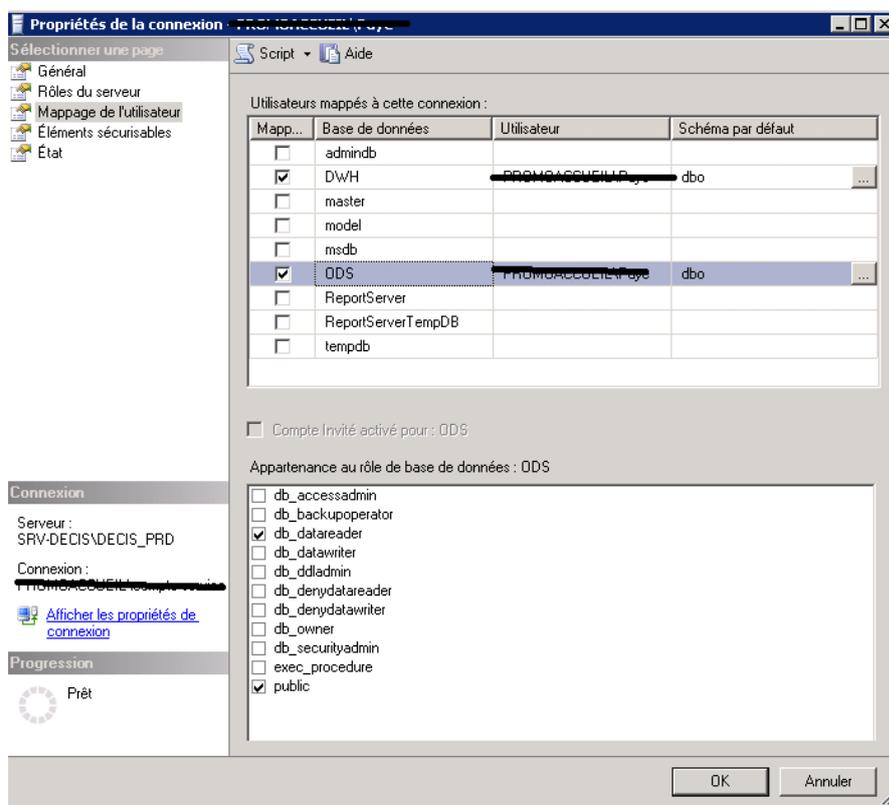
Le système d'information décisionnel a pour objectif de faciliter la mise à disposition des données aux utilisateurs, cependant il convient de gérer les autorisations d'accès aux données des utilisateurs.

#### III.9.3.2 Mise en œuvre

Afin de sécuriser le système décisionnel, l'authentification des utilisateurs via les comptes Windows a été mise en place sur chacun des composants elle permet une sécurisation de bout en bout, base de données, accès au portail affichage des rapports.

#### III.9.3.3 Gestion des droits sur les bases de données

Afin de sécuriser l'accès aux bases de données et à leurs contenus, l'authentification se fait au travers des comptes active directory. L'attribution des droits se fait par la gestion de groupe utilisateurs.

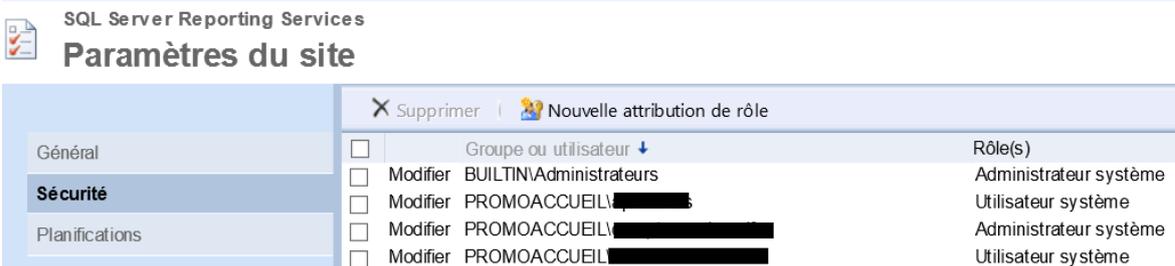


### III.9.3.4 Gestion des droits sur le portail

Afin de sécuriser l'accès aux données, deux niveaux de contrôle d'accès ont été mis en œuvre tout d'abord par la mise en œuvre de droits d'accès au portail avec deux rôles :

- Administrateur système : permet d'afficher et de modifier les attributions de rôles système, les définitions de rôles, les propriétés système et les planifications partagées
- Utilisateur système : permet d'afficher les propriétés système et les planifications partagées et d'utiliser le générateur de rapports.

Les autres utilisateurs se voient refuser l'accès au portail



SQL Server Reporting Services  
**Paramètres du site**

Supprimer | Nouvelle attribution de rôle

<input type="checkbox"/>	Groupe ou utilisateur ↓	Rôle(s)
<input type="checkbox"/>	Modifier BUILTIN\Administrateurs	Administrateur système
<input type="checkbox"/>	Modifier PROMOACCUEIL\██████████	Utilisateur système
<input type="checkbox"/>	Modifier PROMOACCUEIL\██████████	Administrateur système
<input type="checkbox"/>	Modifier PROMOACCUEIL\██████████	Utilisateur système

### III.9.3.5 Gestion des droits sur les rapports

Une arborescence fonctionnelle a été définie sur le portail afin de pouvoir gérer finement les accès aux rapports. Le rôle de chaque groupe d'utilisateurs est déclaré sur chaque dossier du portail.



SQL Server Reporting Services  
**Nouvelle attribution de rôle**

Utilisez cette page pour définir une sécurité basée sur les rôles pour 01 ALFA RETRAITE.

Nom d'utilisateur ou de groupe :

Sélectionnez un ou plusieurs rôles à attribuer au groupe ou à l'utilisateur.

<input type="checkbox"/>	Rôle ↓	Description
<input type="checkbox"/>	Explorateur	Permet d'afficher des dossiers et des rapports, et de s'abonner à des rapports.
<input type="checkbox"/>	Générateur de rapports	Permet de visualiser les définitions de rapports.
<input type="checkbox"/>	Gestionnaire de contenu	Permet de gérer le contenu sur Report Server, notamment des dossiers, des rapports et des ressources.
<input type="checkbox"/>	Mes rapports	Permet de publier des rapports et des rapports liés, de gérer des dossiers, des rapports et des ressources dans le dossier Mes rapports d'un utilisateur.
<input type="checkbox"/>	Serveur de publication	Permet de publier des rapports et des rapports liés sur Report Server.

### **III.10 Exemple de mise en œuvre d'intégration dans l'ODS**

Ce chapitre présente la mise en œuvre du projet décisionnel sur quelques domaines d'EDENIS. J'y décris le déroulement des tâches, les choix et les paramétrages effectués pour chaque composant

#### **III.10.1 Sélection d'un domaine.**

Le service RH met à disposition des établissements des trames Excel afin de générer les contrats de travail remis au salarié. Un contrôle des informations saisies dans ces documents avec les informations enregistrées dans le logiciel de paie est ensuite effectué.

Cette procédure source d'erreur et peu valorisante pour les utilisateurs nécessite d'être revue.

Il a donc été proposé d'intégrer les données des trames Excel afin d'en valider la cohérence avec les données RH

#### **III.10.2 Définition des besoins**

L'existant repose sur une base Excel par établissement listant les contrats générés. Il faut donc intégrer ces informations dans une base de données unique puis mettre en œuvre des rapports listant les cohérences avec les informations du logiciel RH. Le rapport devra permettre un contrôle des saisies faites datant d'une demi-journée au maximum.

#### **III.10.3 Modélisation.**

Les données existantes sont stockées au format XLS dans 18 fichiers possédant des structures similaires.

L'expression des besoins ne fait pas apparaître de besoin d'historisations ou d'analyses dimensionnelles, les données seront donc intégrées dans la base ODS

Une seule table de destination sera créée dans l'ODS à savoir la table CONTRATS dont les colonnes reprennent le format des données sources.

L'ETL effectuera un chargement de chaque fichier sans transformation de format.

### III.10.4 Conception des composants

#### III.10.4.1 L'ETL d'alimentation

L'ETL d'alimentation de la table contrat effectuera les tâches suivantes :

- Supprimer les lignes de la table contrats
- Purger le fichier de LOG
- Effectuer une boucle sur chaque fichier de contrat

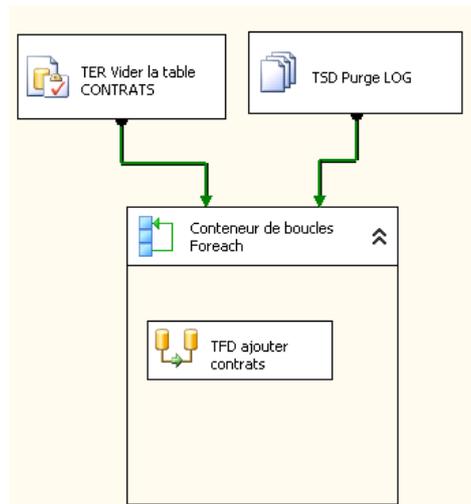


Figure 23 Flux de contrôle Intégration contrats

La boucle effectue les tâches suivantes pour chacun des fichiers :

- Charger toutes les lignes
- Filtrer les lignes vides
- Charger les données dans la table CONTRATS

De plus une gestion d'erreurs charge dans un fichier de LOG

- Les lignes incorrectes du fichier source
- Les lignes dont le chargement dans la table CONTRATS a échoué

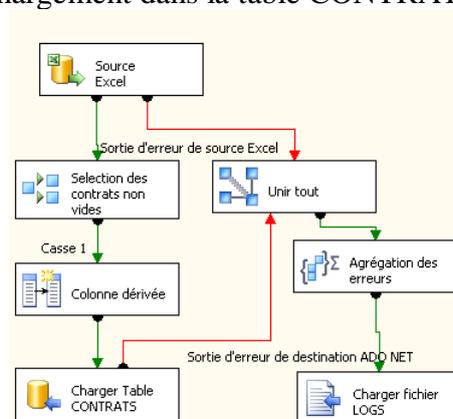


Figure 24 Tâche de flux de données Intégration Contrats

Une planification de l'exécution de l'ETL est programmée par la création d'un travail de l'agent SQL du serveur décisionnel. L'exécution du travail est planifiée toutes les deux

heures entre 07h30 et 21h00. J'ai choisi de définir un délai de moitié inférieur à celui demandé dans les exigences, afin de permettre une intervention du service informatique en cas d'échec d'exécution d'une occurrence tout en respectant les délais métier.

Résumé du fichier journal : Aucun filtre appliqué

Date	ID de l'étape	Serveur	Nom du travail	Nom de l'étape	Notifications
22/04/2016 15:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
22/04/2016 13:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
22/04/2016 11:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
22/04/2016 09:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
22/04/2016 07:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
21/04/2016 19:30:00	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	
21/04/2016 17:30:01	0	SRV-DECIS\DECIS_PRD	MAJ_ODS_CONTRAT_TRAVAIL	(Sortie du travail)	

### III.10.4.2 Le rapport de contrôle

Le rapport de contrôle se connecte sur la source de données ODS et utilise le dataset partagé nommé « contrôle cohérence contrat » qui réalise une jointure entre la table CONTRATS et la table des contrats du logiciel RH nommée OCT\_CONTRAT sur quatre champs : nom prénom, date d'embauche, date de fin de contrat, établissement. Le dataset retourne les lignes en erreur ainsi que le motif de l'erreur.

dated	datef	nom	prenom	niv_cod1	erreur
30/04/2011...	06/05/2011...	[REDACTED]	[REDACTED]	0015	contrat octime sans demande de contrat
30/04/2011...	06/05/2011...	[REDACTED]	[REDACTED]	0005	contrat octime sans demande de contrat
01/05/2011...	08/05/2011...	[REDACTED]	[REDACTED]	0009	contrat octime sans demande de contrat
01/05/2011...	09/05/2011...	[REDACTED]	[REDACTED]	0009	contrat octime sans demande de contrat

Une fois alimenté en données, le rapport final est mis en forme.

Requête paie					
Contrôle contrats Octime/Demande					
Résidence	Nom	Prenom	Date début	datef	Erreur
[niv_co]	[nom]	[prenom]	[dated]	[datef]	[erreur]
	[Count(nom)]				
[&ExecutionTime]					

### **III.11 Exemple de mise en œuvre d'intégration dans le Data warehouse.**

#### **III.11.1 Sélection d'un domaine.**

Le taux de remplissage est l'indicateur de base pour le suivi de l'activité des établissements, son calcul repose sur les informations des séjours des résidents.

#### **III.11.2 Définition des besoins**

Pour ce domaine les besoins suivants ont été exprimés :

- Tout d'abord la perspective d'un changement de logiciel de gestion du résident entraîne la nécessité de stocker les informations de séjours indépendamment du logiciel source.
- Les ETL seront développés sur la base du logiciel de gestion du résident actuel, puis seront modifiés lors de la mise en œuvre du nouveau logiciel.
- Les dates des différents événements d'un séjour devront être stockées en plus des informations du séjour.
- L'actualisation des données sera quotidienne.

Le besoin consiste donc à intégrer les données de séjour dans le Data warehouse avec une modélisation en étoile autour d'une table de fait récapitulative. La mise en œuvre d'une table de fait récapitulative permettra de tracer les événements du cycle de vie du fait métier séjour.

#### **III.11.3 Modélisation.**

La modélisation de l'étoile a été réalisée à l'aide de DBDESIGN qui est un logiciel libre de conception de base de données. Il intègre des fonctionnalités de conception de bases de données, de modélisation, de création et de maintenance dans un environnement homogène.

. Les dimensions suivantes seront associées à la table de fait séjours :

- Temps (dimension conforme existante)
- Etablissement (dimension conforme existante)
- Résident (dimension conforme existante)
- Motif d'entrée
- Motif de sortie

La dimension temps sera associée à chacune des colonnes d'enregistrement des dates des différents évènements du séjour.

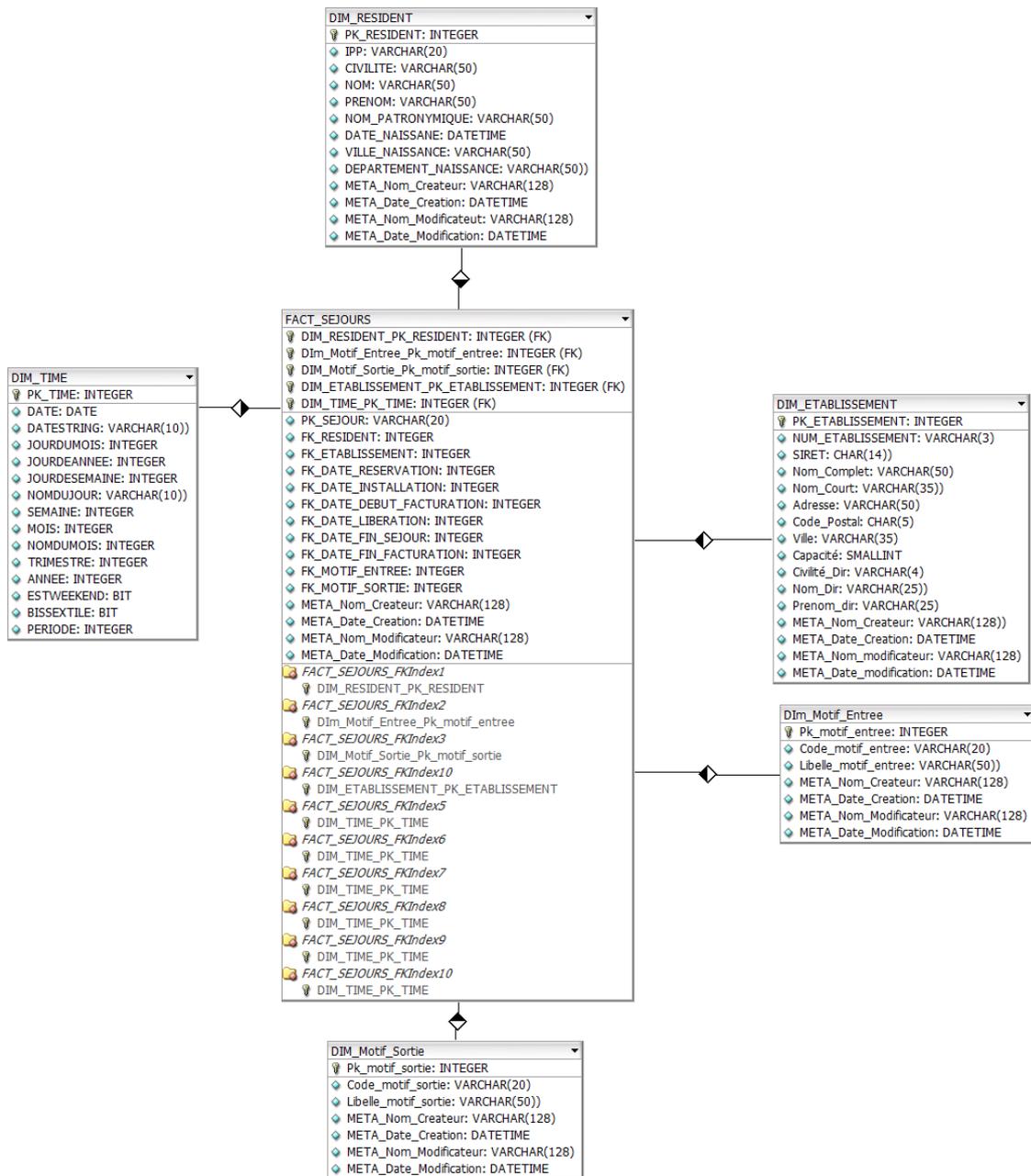


Figure 25 Modélisation en étoile Fait Séjours

L'utilisation d'une table de fait récapitulative a pour conséquence l'utilisation de requête de type Update au niveau des ETL d'alimentation de la table de fait.

### III.11.4 Conception des composants

#### III.11.4.1 La table de fait

A partir de la modélisation de l'étoile, un script de création des tables de dimension et de la table de fait a été généré puis exécuté sur le serveur de base de données. Les liens avec les tables de dimension conformes existantes ont été déclarées ensuite, afin d'aboutir au schéma en étoile ci-dessous (seules les colonnes clefs sont représentées.)

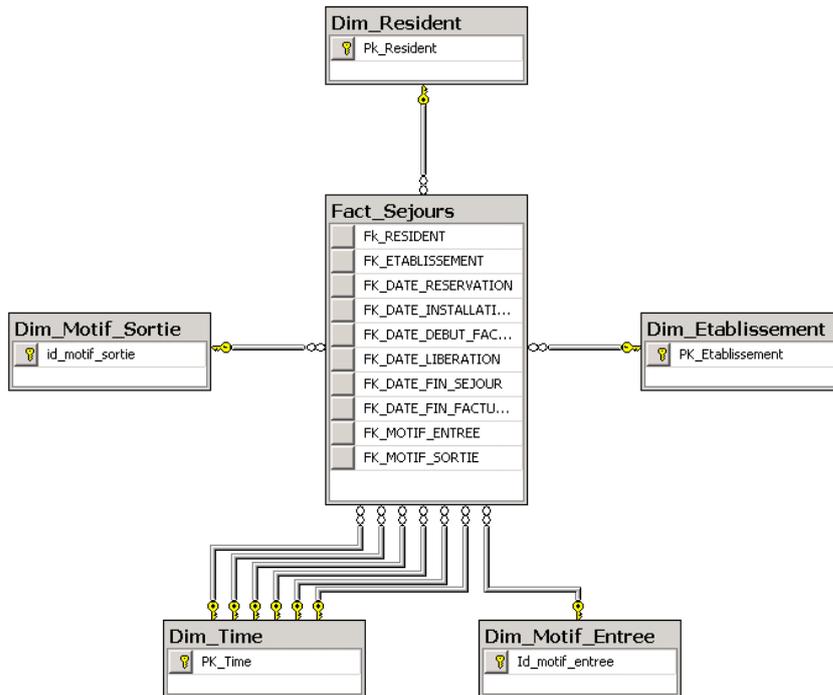


Figure 26 Schéma étoile séjours

### III.11.4.2 Les ETL

Les dimensions TIME et ETABLISSEMENT et RESIDENT sont des dimensions conformes dont les ETL ont été précédemment développés dans le cadre de l'intégration d'autres domaines au Data warehouse. Seul les ETL des dimensions Dim\_Motif\_Sortie et Dim\_Motif\_Entrée seront développés.

Ces deux dimensions de natures similaires seront alimentées par un ETL effectuant les actions suivantes :

1. Extraction de la liste des motifs déclarés dans la base de données métier
2. Agrégation des motifs par code afin de supprimer les doublons
3. Recherche de code motifs non présents dans la table de dimension
4. Ajout des métadonnées
5. Chargement dans la table de dimension.

La table de fait sera alimentée en deux étapes :

- L'insertion des nouveaux séjours où sera ajoutée une ligne contenant la clef métier et les clefs étrangères des dimensions, résidents, établissement

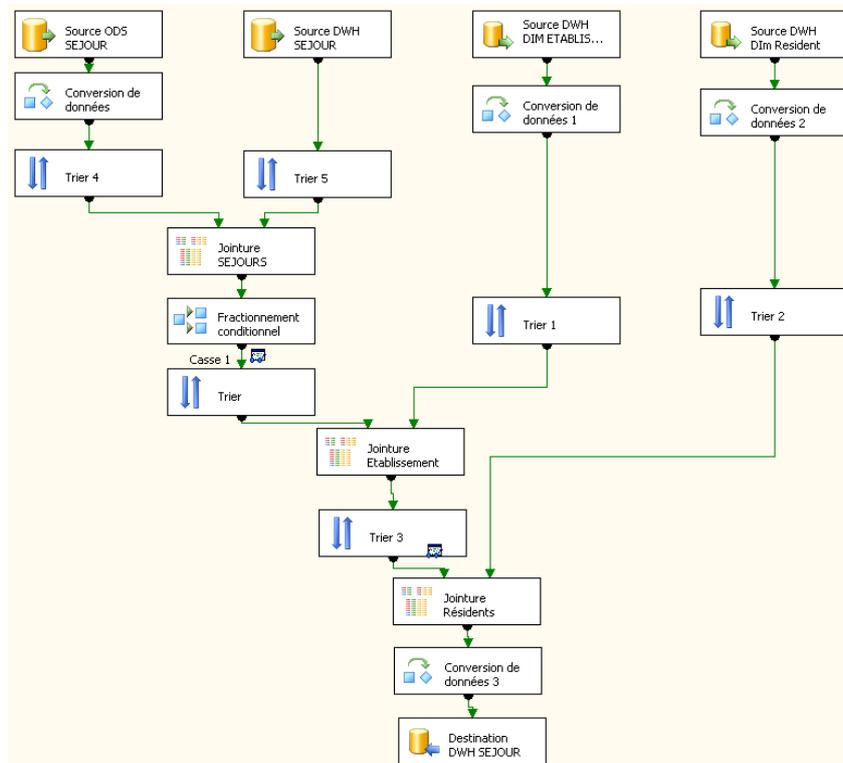


Figure 27 ETL Insertion DWH SEJOURS

- La mise à jour de valeurs d'évènements qui actualisera les valeurs de chacune des colonnes d'évènements non renseignées.

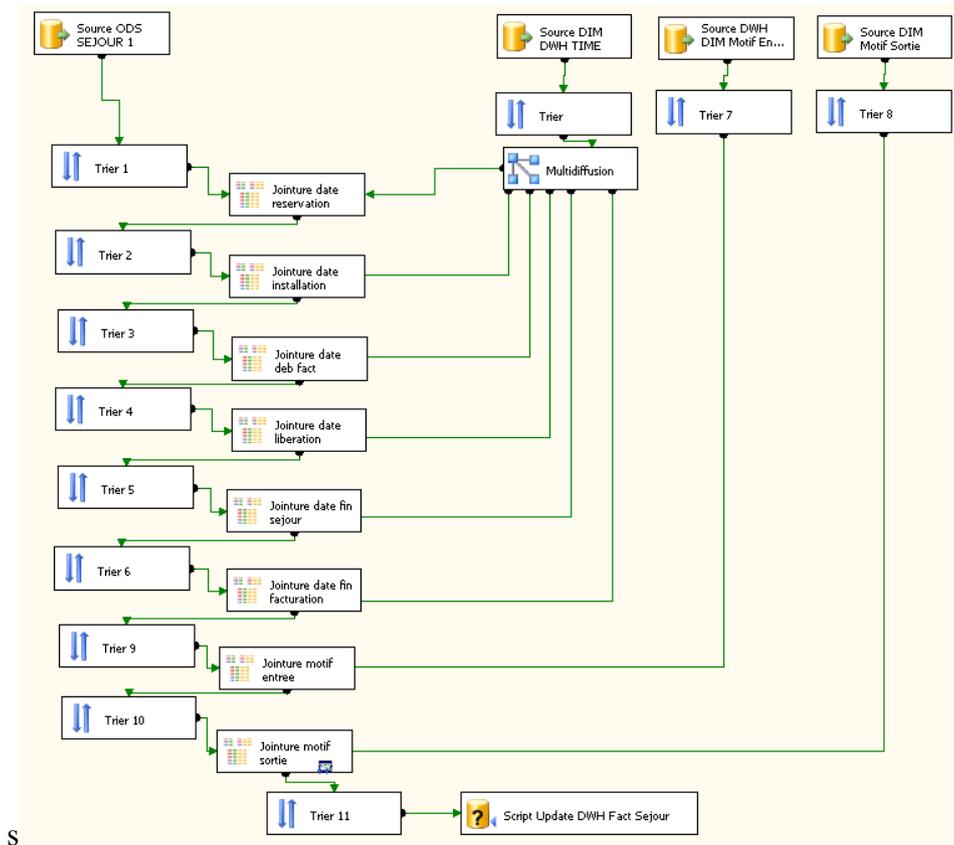


Figure 28 tâche ETL Update Fact Séjour

Les différentes tâches d'executions sont reliées afin d'executer dans un premier temps la mise à jours des deux tables de dimension puis l'execution de l'ajout des nouveaux séjours et enfin la mise à jour des valeurs d'évènements.

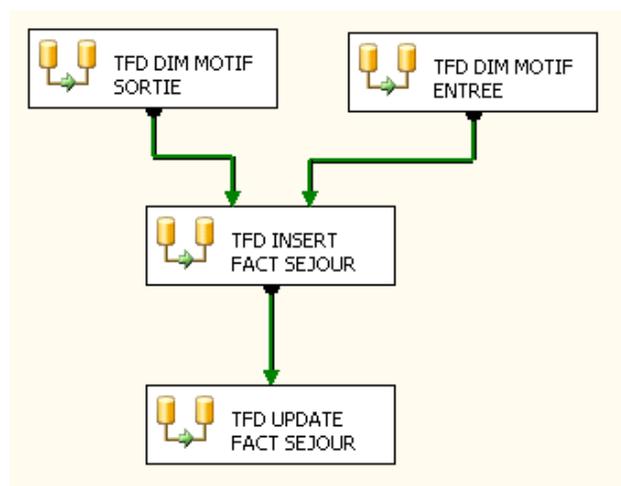


Figure 29 Chainage execution tâches ETL

### **III.12 Evolutions du projet :**

Le projet ayant été mené sur plusieurs mois, il a été impacté par la définition du nouveau schéma directeur des systèmes d'information de l'association. Celui-ci a en effet identifié le besoin de changer la modalité d'acquisition et d'hébergement des logiciels critiques vers le mode SAAS.

#### **III.12.1 Prise en compte des applications SAAS**

Le schéma directeur des systèmes d'informations livré au premier trimestre 2015 a identifié trois domaines fonctionnels comme étant critiques (gestion du résident, gestion des dossiers médicaux, gestion des temps). Les applications supportant ces fonctions seront migrées vers des applications de types SAAS afin de garantir une haute disponibilité.

Les bases de données SAAS ne sont pas accessibles au travers du système d'information d'EDENIS, il est nécessaire de mettre en œuvre avec les éditeurs/hébergeurs des méthodes d'accès sécurisées aux données n'impactant pas la navigation Internet utilisateurs et permettant de respecter les besoins exprimés concernant les fréquences d'actualisation des données. De plus les contraintes légales fortes induites par l'hébergement des données de santé devront être prises en compte.

Le passage au mode SAAS sera l'occasion de changer de progiciel pour la gestion du résident et des dossiers médicaux, cela implique donc de gérer en plus de l'accès aux données le changement de modèle de données.

Les scénarios suivant sont en train d'émerger pour l'intégration des données SAAS.

Un backup des bases de données seront mis à disposition quotidiennement sur un espace sécurisé, chaque ETL devra donc débiter par une phase de transfert de fichier puis de restauration de la base de données sur une instance SQL locale avant de débiter les tâches d'intégration.

Pour les bases de données de santé nominative les données seront anonymisées avant la mise à disposition afin de répondre aux exigences réglementaires. La dimension résident ne sera donc pas exploitable sur ce domaine.

### III.12.2 Intégration aux outils bureautique SAAS.

A partir du mois de juin 2015 EDENIS a commencé à déployer auprès de ses utilisateurs la suite bureautique O365 Entreprise. Il s'agit de la version SAAS de Microsoft Office proposant un ensemble de service cloud et de licence pour des installations locales sur les postes de travail.

Ce déploiement a permis aux utilisateurs de disposer des dernières versions d'Excel incluant notamment les nouveaux composants de business intelligence :

- Power Query

Microsoft Power Query pour Excel est un complément de collecte de données. Il permet d'intégrer des données provenant de multiples sources (bases de données relationnelles, Excel, fichiers texte et fichiers XML, flux OData, pages Web, Hadoop HDFS, etc.).

Il assure des fonctions d'ETL : Import, fusion, mise en forme, nettoyage des données

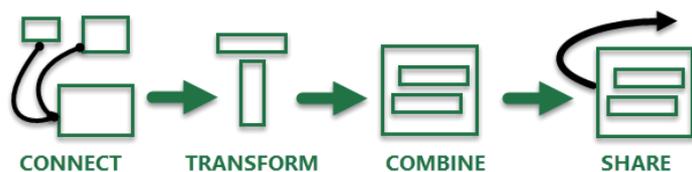


Figure 30 Fonctionnalités Power Query

- Power Pivot

Microsoft Power Pivot pour Excel est un complément de diffusion de données permettant de définir des relations entre plusieurs sources de données. Les données ainsi intégrées peuvent ensuite être mises à disposition au travers d'un modèle de données dans des outils d'analyse.

Le langage DAX (Data Analysis Expressions) est un langage de formule qui permet aux utilisateurs de définir des calculs personnalisés dans les tables PowerPivot (colonnes calculées) et dans les tableaux croisés dynamiques Excel (mesures).

- Power Map

Microsoft Power Map pour Excel est un complément d'analyse des données permettant de tracer des données géographiques et temporelles sur un globe ou une carte 3D personnalisée.

- Power View

Microsoft Power View pour Excel est un complément d'analyse des données permettant la création et la mise à disposition de tableaux de bord incluant de nombreux types de graphiques dynamiques.

- Power BI

De plus pour certains utilisateurs avancés (contrôle de gestion, service informatique, direction financière) l'association a acquis des licences Office 365 Power BI.

Power BI est une suite d'outils SaaS permettant d'analyser des données et de partager des informations. Elle permet de créer et de partager en ligne des tableaux de bord ,des rapports et des jeux de données

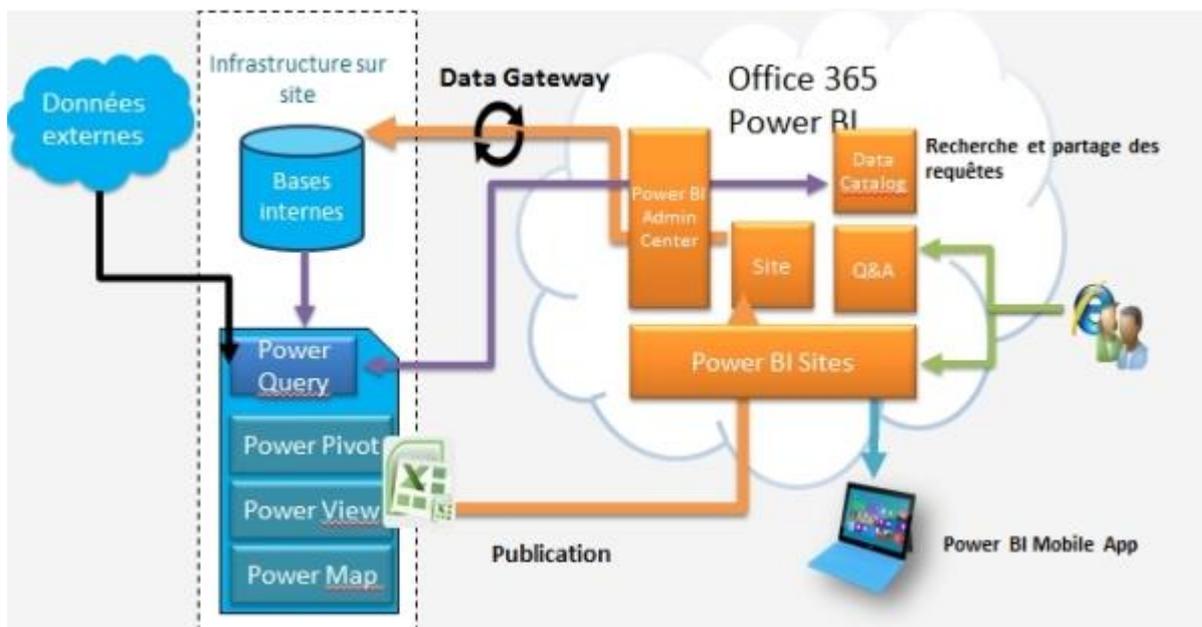


Figure 31 Microsoft Composant Power BI

Tous ces composants dit de BI personnelle favorisent l'autonomie des utilisateurs dans la production et l'exploitation de données BI. Ils complètent les fonctionnalités du portail SSRS en permettant une décentralisation de l'information par l'exploitation directe des datasets et source de données sur les postes de travail. La mise à disposition d'un portail cloud personnel facilitera également l'accès aux utilisateurs nomades.

Ces composants sont installés chez un nombre restreint d'utilisateurs, leurs diffusions à un public plus large nécessitera un fort accompagnement au changement, car en favorisant l'autonomie ils entraînent un besoin en formation plus important des utilisateurs.

### **III.12.3 Evolution du système d'information**

Comme je l'ai évoqué dans la présentation du système d'information, le siège héberge actuellement deux clusters de virtualisation, un cluster CITRIX d'ancienne génération et un cluster HYPER-V comprenant une ferme RDS 2012 Server mis en œuvre en janvier 2016. Les applications métiers sont progressivement migrées sur le cluster HYPER-V.

Le SID devant évoluer en parallèle du SIG, tous les composants décisionnels seront également migrés sur ce nouvel environnement.

Cette migration sera un projet au moins aussi complexe que la migration d'un logiciel métier du fait du nombre de composants à prendre en compte, et surtout de la nécessité de contrôler le bon fonctionnement de chacun des ETL notamment au niveau des connecteurs aux sources de données.

Le système d'information de gestion fait l'objet d'un rattrapage en termes de version d'application et d'OS afin de profiter des mises à jour de sécurité et fonctionnelles disponibles, le système d'information décisionnel doit également en faire l'objet

La version de Microsoft SQL server utilisée est la version 2008 R2, ce qui correspond à 2 version majeure de retard sur la dernière version disponible sur le marché sans prendre en compte la sortie d'une nouvelle version prévue d'ici la fin de l'année

Le support principal de SQL server 2008 R2 a été stoppé le 08 juillet 2014 seules des mises à jour critiques seront réalisé pour ce produit jusqu'au 09 juillet 2019. Afin de ne pas maintenir une plateforme obsolète sa mise à jour devra être réalisée prochaine.

### **III.12.4 Autres perspectives**

La création du système d'information décisionnel a été l'occasion pour l'association de s'interroger sur sa gouvernance des données et de mettre en œuvre des bonnes pratiques afin de l'améliorer. Cependant il s'agit d'un domaine où l'amélioration reste possible au bénéfice de tous les acteurs.

## Conclusion

En 2016 le système d'information opérationnel de l'association va fortement évoluer sous l'impulsion du SDSI et le SID, débuté il y a maintenant plus de trois ans, devra évoluer en parallèle. C'est le moment de tirer un bilan de cette première phase.

### **Bilan de la proposition projet :**

Lors de l'étude du pré projet j'avais basé ma proposition sur quatre arguments, trois ans après on peut en tirer les résultats suivants :

- Ce type de projet nécessite une connaissance approfondis des processus EDENIS

Le fait que le système d'information décisionnel soit construit et exploité par les mêmes personnes qui conçoivent et exploite le système d'information opérationnel facilite fortement le déroulement du projet décisionnel.

Les informaticiens connaissent déjà les processus métier et le vocabulaire associé, il est plus facile aux utilisateurs métier de dialoguer avec des personnes qui travaillent avec eux depuis des années sur ces processus.

- Il est nécessaire de disposer de ressources compétentes et rapidement mobilisables pour réaliser les évolutions nécessaires.

L'association dispose désormais en interne de compétences pour l'exploitation et l'évolution de son système d'information décisionnel. Elle n'est pas dépendante de la disponibilité d'un budget pour une intervention d'une Société de Conseil en Ingénierie Informatique (SCII) qui ne peut garantir l'affectation d'un intervenant connaissant l'existant du projet.

- Le système d'information étant entièrement basé sur un environnement Microsoft, il paraît cohérent d'utiliser une solution décisionnelle du même éditeur.

Le choix de la plateforme décisionnelle n'a pas fait l'objet d'une étude approfondie comme recommandé dans les méthodes de gestion de projet. Cependant le bon retour des utilisateurs sur ces outils lors des projets précédents et la validation de la couverture fonctionnelle de ce produit lors de la phase de pré projet, a permis de démarrer le projet sereinement. A ce jour il ne fait aucun doute sur la pertinence de ce choix, le SDSI ayant validé la prévalence des solutions Microsoft dans le SIG et la poursuite du projet décisionnel.

- Les investissements fait lors des tests de la solution à savoir l'acquisition des licences et la formation de membres du service informatique et du contrôle de gestion pourront être réutilisé dans le cadre de ce nouveau projet.

Le projet a été entièrement mené à partir des investissements des projets précédents, aucune acquisition n'a été faite ces trois dernières années. Les investissements prévus sont ceux nécessaire à tous système d'information à savoir acquérir les licences permettant l'actualisation des versions. Les autres investissements tels que ceux concernant la plateforme d'exécution, la sauvegarde, ... sont mutualisés avec ceux du système d'information de gestion, l'impact du décisionnel étant faible.

### **Bilan des objectifs du projet.**

Trois ans après il s'avère que la quasi-totalité des rapports et des tableaux de bords sont réalisés à partir du système d'information décisionnel. Le service contrôle de gestion est autonome dans la création et la diffusion des rapports.

Les utilisateurs du siège et des résidences ont l'habitude de recevoir dans leurs boites mails des rapports automatiques. Des demandes de mise à disposition sont même désormais remonté au service informatique par le biais de la plateforme d'helpdesk signe d'une bonne perception de l'outil par les opérateurs métiers.

Toutefois deux points ne sont pas encore finalisés :

L'intégration des données de soins du résident a été reporté dans l'attente du changement de logiciel. Celui est en cours de déploiement et l'intégration des données devrait avoir lieu cette année.

La plupart des rapports demandé par les utilisateurs ne nécessite pas les fonctionnalités du Data warehouse et sont donc alimenté par l'ODS. L'évolution des applications métier devrait corriger ce phénomène de sous-utilisation du Data warehouse puisque lui seul contiendra les données antérieures au changement de logiciel et permettra la continuité de la fourniture des rapports.

### **Bilan Humain**

Le choix de réaliser ce projet sans aucun prestataire extérieur a été très valorisant pour tous les acteurs Ce projet a nécessité un investissement conséquent afin de réussir sa mise en

œuvre en parallèle de mes fonctions courantes. L'investissement et la forte implication du service contrôle de gestion a également été un facteur de succès de ce projet.

La création du système d'information décisionnel a également permis d'effectuer un rapprochement des équipes SI et métier, en obligeant les informaticiens à s'intéresser au fonctionnement quotidien des services métier et non plus seulement à leurs outils.



## Bibliographie

- BRESLIN, M. (2004). Data warehousing battle of the giants: Comparing the basics of the kimball and Inmon Models. *Bussiness Intelligence Journal*.
- GOEDE, R., & HUISMAN, M. (2010). The suitability of agile systems development methodologies for data warehouse development. *The proceeding of the international Conference on information management an evaluation*, (pp. 99-106). North west university, South Africa.
- INMON, B. (1990). *Building the Data WAREHOUSE*.
- INMON, W. (2002). *Building the Data Warehouse. 3rd Edition*. John Wiley & sons.
- KIMBALL, R. (2004, Décembre 4). *Intelligent Enterprise*.
- KIMBALL, R., & REEVES, L. (2000). *Concevoir et déployer un datawarehouse*. Eyrolles.
- KIMBALL, R., & ROSS, M. (2013). *THE Data Warehouse Toolkit : The Definitive guide to dimensional modeling 3rd Edition*. WILEY.
- KIMBALL, R., & ROSS, M. (2016). *The Kimball Group Reader, Remastered Collection*. Wiley.
- KIMBALL, R., ROSS, M., Thornthwaite, W., MUNDY, J., & BECKER, B. (2008). *The Data Warehouse Lifecycle Toolkit*. Wiley Publishing, second edition.
- Kumar, A. (2015, Aout 15). *top-10-methods-to-improve-etl-performance-using-ssis*. Récupéré sur developer.com: <http://www.developer.com/db/top-10-methods-to-improve-etl-performance-using-ssis.html>
- RIVARD, E. (2012). Proposition d'une méthodologie agile en intelligence d'affaires pour réduire les risques d'échecs. *Essai pour l'obtention du grade de maître en technologies de l'information*. QUEBEC: Université de SHERBROOKE.
- TESTE, O. (2000). Modélisation et manipulation d'entrepôts de données complexes et historisées. *Thèse doctorat informatique Université Paul Sabatier*. TOULOUSE.

## Table des annexes1

Annexe 1 Liste des établissements .....	93
Annexe 2 Taux d'occupation .....	94
Annexe 3 Configuration source de données.....	95
Annexe 4 Projets identifiés par le SDSI.....	96
Annexe 5 Script création dimension temps.....	97
Annexe 6 Script SQL : Afficher les 50 dernières requêtes exécutées et exemple de résultat .....	98
Annexe 7 Script VBA : Alimenter une base de données Excel à partir d'un fichier contenant des objets et des cellules nommées.....	99
Annexe 8 Gartner Magic Quadrant for Business Intelligence and Analytics Platforms 2016 .....	101

## Annexe 1 Liste des établissements

<p><b>1</b> <b>TOULOUSE</b> <b>CAROLINE BARON *</b> 51 Avenue Jean Rieux 31500 TOULOUSE 05 34 66 72 00 accueilcarolinebaron@edenis.fr</p> <p><b>2</b> <b>LA PASTELLIÈRE</b> 36 Boulevard Jean Brunhes 31300 TOULOUSE 05 34 51 28 88 accueilpastelliere@edenis.fr</p> <p><b>3</b> <b>LA COTONNIÈRE *</b> 29 Chemin de la Garonne 31200 TOULOUSE 05 62 72 03 27 accueilcotonniere@edenis.fr</p> <p><b>4</b> <b>BLAGNAC</b> <b>EMERAUDE</b> 14 Rue des Amandiers 05 61 71 44 50 accueillemeraude@edenis.fr</p> <p><b>5</b> <b>BEAUZELLE</b> <b>EDELWEISS *</b> 5 Allée du Plantaurel 05 62 74 51 00 accueilleedelweiss@edenis.fr</p> <p><b>6</b> <b>SAINT-ALBAN</b> <b>LE CLOS DES AMANDIERS *</b> 27 Rue Bernard Amiel 05 34 42 14 14 accueillamandiers@edenis.fr</p> <p><b>7</b> <b>CASTELGINEST</b> <b>LA TOUR TOTIER</b> 9 Route de Pechbonnieu 05 61 70 45 18 accueilltotier@edenis.fr</p> <p><b>8</b> <b>PIBRAC</b> <b>LA HOULETTE *</b> 3 Rue Albert Camus 05 61 86 44 99 accueillhoulette@edenis.fr</p> <p><b>9</b> <b>PLAISANCE-DU-TOUCH</b> <b>LE PRAT</b> Avenue de Lingfield 05 61 07 37 37 accueilprat@edenis.fr</p>	<p><b>10</b> <b>TOURNEFEUILLE</b> <b>LE GRAND MARQUISAT *</b> 28 Rue Berthelot 05 62 48 20 30 accueilgrandmarquisat@edenis.fr</p> <p><b>11</b> <b>ESCALQUENS</b> <b>BORDE-HAUTE</b> 38 bis Avenue Borde-Haute 05 61 75 30 00 accueilbordehaute@edenis.fr</p> <p><b>12</b> <b>PORTET-SUR-GARONNE</b> <b>L'AUTA</b> 3 bis Rue du Port 05 61 72 30 60 accueillauta@edenis.fr</p> <p><b>13</b> <b>VILLENEUVE-TOLOSANE</b> <b>LE PIN</b> 2 Rue René Cassin 05 62 87 28 87 accueilpin@edenis.fr</p>	<p><b>14</b> <b>MURET</b> <b>MARIE-ANTOINETTE *</b> 6 Boulevard Aristide Briand 05 61 51 28 97 accueilantoinette@edenis.fr</p> <p><b>15</b> <b>LE BARRY</b> 18 Rue des Marins 05 61 56 45 15 accueilbarry@edenis.fr</p> <p><b>16</b> <b>SAINT-GAUDENS</b> <b>LE MAS SAINT PIERRE</b> 24 Avenue de Boulogne 05 62 00 35 34 accueilmas@edenis.fr</p> <p><b>17</b> <b>L'ENSOLEILLADE</b> 5 Rue de la Résidence 05 62 00 92 85 accueilleSOLEILLADE@edenis.fr</p> <p><b>18</b> <b>MONTAUBAN</b> <b>LES SAULES *</b> 12 Rue des Saules 05 63 26 72 67 accueilleSAULES@edenis.fr</p>
--	---	---

**\* EHPAD équipé d'une unité de vie protégée**

Siège social 05 62 27 57 27  
3, rue Claude-Marie Perroud BP 10647  
31106 TOULOUSE CEDEX 1

**N° Vert 0 805 805 901**

Association à but non lucratif (loi 1901)

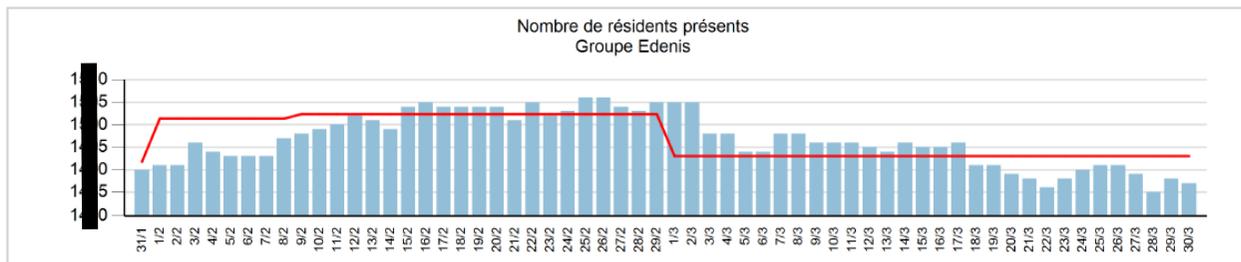
**www.edenis.fr**  
GROUPE **edenis**  
RÉSIDENCES MÉDICALISÉES SENIORS

## Annexe 2 Taux d'occupation

Taux d'occupation exploitation

Projection du mois de

Résidences	Lits	Réalisés				Objectif 2016	mars		Capacité restante	Nbre résid	Sorties		Variation du jour	Entrées Prév 15j			
		2012	2013	2014	2015		Objectif	TO			Totales	Décès					
01	Auta	85	4,7%	0,5%	0,0%	0,1%	0,95%	0,06%	0,17%			-5	5	4	-1	1	84,0%
02	Tour totier	90	9,3%	4,2%	2,3%	2,4%	0,71%	1,44%	5,41%			0	2	2	0	0	62,4%
03	Marie antoinette	100	6,4%	1,1%	0,6%	0,2%	0,49%	0,50%	7,55%			-1	5	3	0	2	81,6%
04	Prat	93	7,0%	0,3%	0,1%	0,2%	0,92%	0,77%	9,38%			-2	3	3	0	0	67,4%
05	Barry	87	6,9%	1,3%	0,2%	0,8%	0,38%	0,83%	9,41%			1	0	0	0	2	75,9%
06	Houlette	80	4,0%	0,5%	0,1%	0,9%	0,27%	0,00%	5,85%			1	4	2	0	0	70,1%
07	Pastelliere	88	4,1%	0,1%	0,2%	0,7%	0,62%	0,48%	7,98%			-4	7	2	0	0	79,0%
08	Cotonniere	80	2,9%	0,0%	0,5%	0,3%	0,77%	0,38%	6,09%			1	2	2	0	0	77,9%
09	Pin	88	7,2%	0,0%	0,2%	0,4%	0,64%	0,89%	4,43%			-3	6	4	0	1	81,5%
11	Ensoleilade	88	9,0%	2,4%	0,1%	0,5%	0,75%	0,73%	0,51%			-5	8	5	0	0	67,5%
12	Mas st pierre	82	1,1%	0,4%	0,3%	0,0%	0,34%	0,34%	2,76%			-1	4	1	0	1	64,5%
13	Gd marquisat	80	6,5%	0,7%	0,5%	0,9%	0,26%	0,63%	8,19%			-1	1	1	0	0	82,1%
14	Caroline baron	80	8,6%	0,4%	0,1%	0,3%	0,14%	0,44%	7,70%			-1	4	4	0	0	71,8%
16	Borde haute	80	5,0%	0,4%	0,0%	0,1%	0,12%	0,88%	6,29%			0	3	1	0	0	80,5%
17	Emeraude	90	3,5%	0,5%	0,7%	0,6%	0,14%	0,11%	2,01%			-1	4	2	0	2	68,7%
18	Clos des amandiers	80	0,2%	0,5%	0,9%	0,7%	0,40%	0,25%	5,40%			0	4	2	0	0	88,3%
19	Saules	105	1,3%	0,5%	0,1%	0,2%	0,66%	0,86%	9,45%			2	1	1	0	0	100,0%
20	Edelweiss	80	1,9%	0,6%	0,8%	0,1%	0,28%	0,50%	6,13%			2	4	4	0	0	100,0%
<b>Groupe Edenis</b>		<b>1 556</b>										<b>-17</b>	<b>67</b>	<b>43</b>	<b>-1</b>	<b>9</b>	<b>78,2%</b>



31/03/2016 10:31:26

## Annexe 3

### Configuration source de données

Dossier racine > SHARED\_DATA\_SOURCES Dossier racine | Mes abonnements | Paramètres du site | Aide

**EDENIS Reporting BI**  
**BDD\_DWH\_REPORTING** Rechercher 

 Supprimer |  Déplacer |  Générer le modèle

**Propriétés**  
Abonnements  
Éléments dépendants  
Sécurité

Nom :

Description :

Masquer en mode Mosaïque  
 Activer cette source de données

Type de source de données :

Chaîne de connexion :

Se connecter avec :

Informations d'identification fournies par l'utilisateur qui exécute le rapport  
Afficher le texte suivant pour demander à l'utilisateur un nom d'utilisateur et un mot de passe :  
  
 Utiliser comme informations d'identification Windows lors de la connexion à la source de données

Informations d'identification stockées en sécurité dans le serveur de rapports

Nom d'utilisateur :

Mot de passe :

Utiliser comme informations d'identification Windows lors de la connexion à la source de données  
 Emprunter l'identité de l'utilisateur authentifié une fois la connexion établie à la source de données

Sécurité intégrée de Windows  
 Informations d'identification non requises

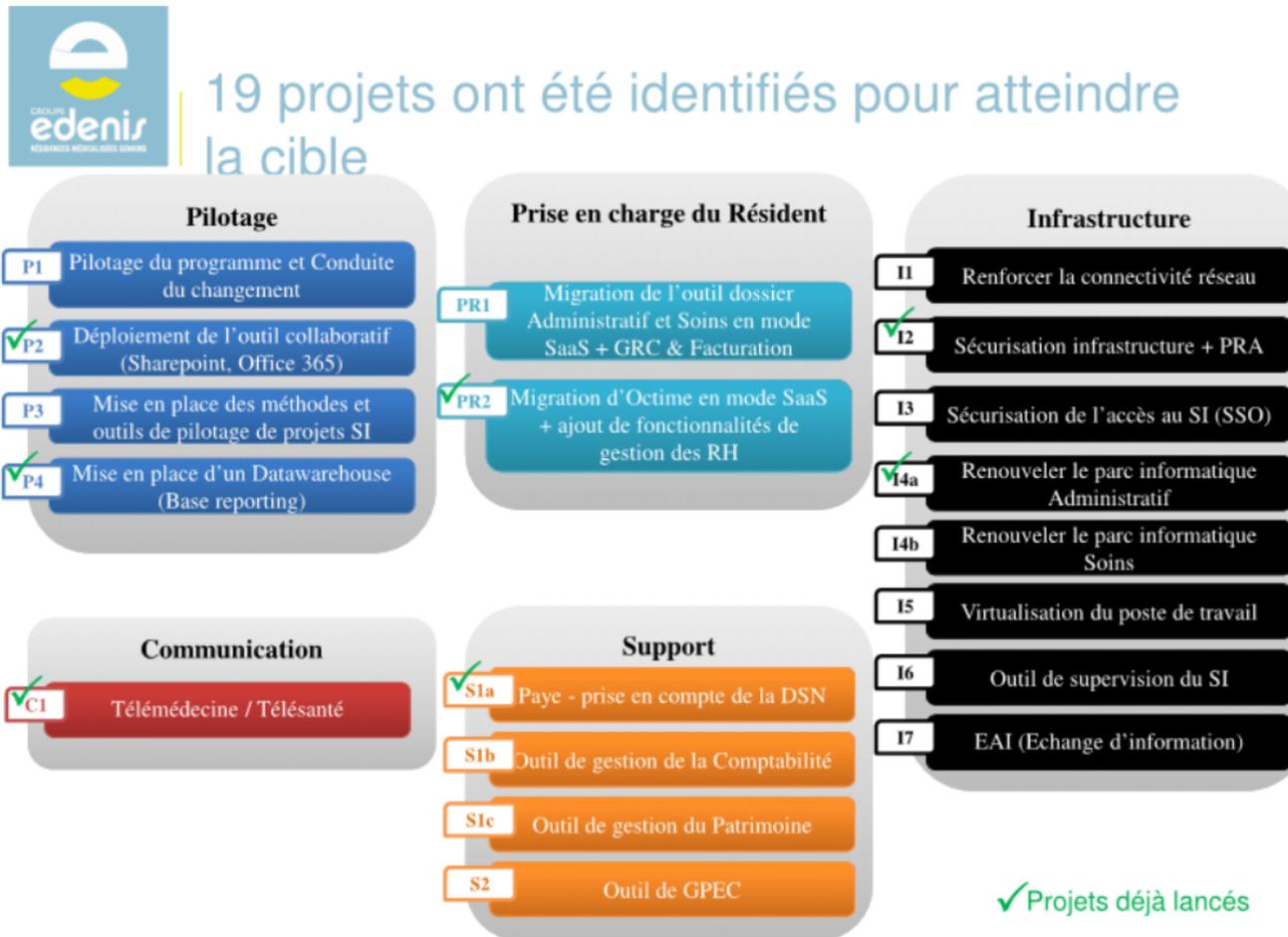
Comment mettre une page en format « paysage » ?

Positionnez le curseur sur la page et vérifiez que le saut est bien un « saut de section (page suivante) ».

Allez dans « Fichier », « Mise en page ». Sous l'onglet « Marges », cliquez dans la rubrique « Orientation » sur « Paysage ».

Puis dans la rubrique « Aperçu » du même onglet, sélectionnez « Appliquez à » : « cette section » Cliquez sur OK.

## Annexe 4 Projets identifiés par le SDSI



## Annexe 5

### Script création dimension temps

```

-- Création d'une table de dimension temps
-- Chaque enregistrement correspond à une date
-- Paramètres : @startdate et @enddate
-- les noms de colonnes sont à personnaliser

DROP TABLE TimeDimension
GO
CREATE TABLE TimeDimension
(
    DateID                DATE                NOT NULL PRIMARY KEY,
    MonthName              VARCHAR(9)         NOT NULL,
    MonthName_FR           VARCHAR(9),
    MonthNumber            INT                NOT NULL,
    DayNumber              INT                NOT NULL,
    DayName                VARCHAR(9)         NOT NULL,
    DayName_FR             VARCHAR(8),
    Quarter                INT                NOT NULL,
    QuarterName            CHAR(7)           NOT NULL,
    Year                   INT                NOT NULL
)
GO
BEGIN
    SET DATEFIRST 1
    DECLARE @startdate DATETIME
    DECLARE @enddate  DATETIME
    DECLARE @date     DATETIME
    DECLARE @offset   INT
    DECLARE @         dateId      DATE
    SET @startdate = '2010-01-01'
    SET @enddate   = '2020-12-31'
    SET @offset    = 0
    SET @date      = DATEADD(dd, @offset, @startdate)
    WHILE @date <= @enddate
    BEGIN
        SET @dateId = @date

        INSERT INTO
        TimeDimension(DateID,MonthName,MonthNumber,DayNumber,DayName,Quarter,QuarterName,Year)
        VALUES (@         dateId,      --DateID
                DATENAME(mm, @date),   --MonthName
                DATEPART(mm, @date),   --MonthNumber
                DATEPART(dd, @date),    --DayOfMonth
                DATENAME(DW, @date),    --DayName
                DATEPART(qq, @date),    --Quarter
                'Q' + DATENAME(qq, @date), --QuarterName
                DATEPART(yy, @date)    --Year
        )
        SET @offset = @offset + 1
        SET @date   = DATEADD(dd, @offset, @startdate)
    END
END
GO

-- Conversions des noms de jours et de mois en Français
SET LANGUAGE French
UPDATE TimeDimension SET DayName_FR = DATENAME(DW, DateId), MonthName_FR = DATENAME(mm, DateId)
SET LANGUAGE US_English

```

## Annexe 6

### Script SQL : Afficher les 50 dernières requêtes exécutées et exemple de résultat

```

SELECT TOP 50 creation_time,
total_worker_time/execution_count AS [Avg CPU Time],
SUBSTRING(st.text, (qs.statement_start_offset/2)+1,((CASE qs.statement_end_offset
WHEN -1 THEN DATALENGTH(st.text)
ELSE qs.statement_end_offset END - qs.statement_start_offset)/2) + 1) AS statement_text
FROM sys.dm_exec_query_stats AS qs
CROSS APPLY sys.dm_exec_sql_text(qs.sql_handle) AS st
ORDER BY creation_time DESC;

```

	creation_time	Avg CPU Time	statement_text
1	2016-05-04 16:12:11.977	11874	SELECT * FROM ALFA_EXP_OSIRIS_PAT_ADRESSES WHERE IPP_PATIENT = '005475' AND BL_PRINCIPALE = -1
2	2016-05-04 16:12:09.633	11129	SELECT * FROM ALFA_EXP_OSIRIS_PAT_CONTACTS WHERE IPP_PATIENT = '005486'
3	2016-05-04 16:12:09.587	11227	SELECT * FROM ALFA_EXP_OSIRIS_PAT_COMPLEMENTAIRES WHERE IPP_PATIENT = '005486' ORDER BY DROITS_OUVERTS_AU DESC
4	2016-05-04 16:12:09.540	11158	SELECT * FROM ALFA_EXP_OSIRIS_PAT_COMPLEMENTAIRES WHERE DROITS_OUVERTS_AU IS NULL AND IPP_PATIENT = '005486'
5	2016-05-04 16:12:09.493	16367	SELECT * FROM ALFA_EXP_OSIRIS_PAT_COUVERTURE WHERE IPP_PATIENT = '005486' ORDER BY DROITS_OUVERTS_DU DESC
6	2016-05-04 16:12:09.353	29602	SELECT * FROM ALFA_EXP_OSIRIS_PAT_TELECOMS WHERE IPP_PATIENT = '005486'
7	2016-05-04 16:12:09.290	10973	SELECT * FROM ALFA_EXP_OSIRIS_PAT_ADRESSES WHERE IPP_PATIENT = '005486'
8	2016-05-04 16:12:09.040	10786	SELECT * FROM ALFA_EXP_OSIRIS_PAT_ADRESSES WHERE IPP_PATIENT = '005486' AND BL_PRINCIPALE = -1
9	2016-05-04 16:12:08.120	9748	SELECT * FROM ALFA_EXP_OSIRIS_SEJOURS WHERE ((DATE_ENTREE >= '03-05-2016' AND (DATE_SORTIE < '04-05-2016' OR (DATE_SORTIE = '04-05-2016' AN...
10	2016-05-04 16:12:06.057	7585	SELECT * FROM ALFA_EXP_OSIRIS_PAT_CONTACTS WHERE IPP_PATIENT = '005480'

## Annexe 7

### Script VBA : Alimenter une base de données Excel à partir d'un fichier contenant des objets et des cellules nommées.

```
Sub ajout_bdd()
Dim BDD_NAME As String
Dim BDD As String
Dim compteur As Integer
Dim formulaire As String
Dim j As Integer
Dim valeur As String
Dim Residence As String
Residence = Feuill.ComboBox1.Value
formulaire = ThisWorkbook.Name
compteur = 2
BDD_NAME = "Edd.xls"
Select Case Residence
Case "RESIDENCE L'AUTA"
valeur_residence = "01. AUTA"
' suppression des autres cas pour la mise en page...
Case Else
MsgBox "RESIDENCE INCONNUE"
End Select
BDD = "\\srv-fich02\VERS SIEGE\03 Direction Financière\04 PAIE\Echanges Externes\Contrats\" & valeur_residence &
"\Réservé RH\BDD\"
'ouverture de la base de donnée
Workbooks.Open (BDD & BDD_NAME)
compteur = Range("A65536").End(xlUp).Row + 1
' ajout des données objets
Workbooks(formulaire).Activate
For Each Obj In ActiveSheet.OLEObjects
    For j = 1 To 50
Workbooks(BDD_NAME).Activate
        If Obj.Name = ActiveSheet.Cells(1, j).Value Then
            ActiveSheet.Cells(compteur, j).Value = Obj.Object.Value
            j = 240
        End If
    If j = 50 Then
        End If
    Next j
    Workbooks(formulaire).Activate
Next Obj
' ajout des données cellules nommées
Workbooks(formulaire).Activate
Set nms = ActiveWorkbook.Names
For Each cellule In ThisWorkbook.Sheets(1).Range("A1:J50").Cells
    For I = 1 To nms.Count
        MsgBox (nms(I).Name)
        If nms(I).RefersToRange.Address = cellule.Address Then
            ' MsgBox (nms(I).Name)
            Workbooks(BDD_NAME).Activate
            For K = 1 To 200
                If nms(I).Name = ActiveSheet.Cells(1, K).Value Then
                    ActiveSheet.Cells(compteur, K).Value = cellule.Value
                    K = 250
                    I = nms.Count
                    End If
                If K = 250 Then
                    ' MsgBox (cellule.Name.Name)
                    End If
            Next '
        End If
    Next I
Next
Next
```

```
'fermeture et sauvegarde la BDD  
Workbooks(BDD_NAME).Close True  
'verouiller protection  
Workbooks(formulaire).Activate  
ThisWorkbook.Protect  
End Sub
```

## Annexe 8

# Gartner Magic Quadrant for Business Intelligence and Analytics Platforms 2016



## Liste des figures

Figure 1 Organigramme Direction générale EDENIS .....	9
Figure 2 Organigramme EHPAD .....	10
Figure 3 Organigramme service informatique .....	12
Figure 4 Schéma infrastructure EDENIS .....	13
Figure 5 Cartographie applicative .....	14
Figure 6 Répartition hébergement SDSI.....	14
Figure 7 Gestion de la qualité des données .....	27
Figure 8 Le cycle de vie d'un projet décisionnel (KIMBALL & ROSS, THE Data Warehouse Toolkit : The Definitive guide to dimensional modeling 3rd Edition, 2013) ..	29
Figure 9 Le cycle de vie d'un projet décisionnel selon la méthode Inmon (INMON W. , 2002) .....	30
Figure 10: Schéma base de données gestion du résident .....	42
Figure 11 : Diagramme de Gantt Projet EDENIS .....	47
Figure 12 Diagramme phases projet .....	48
Figure 13GPO diffusion raccourci portail décisionnel .....	61
Figure 14Création d'une source de données .....	62
Figure 15 Création de dataset.....	63
Figure 16:Commandes TELNET test serveur SMTP .....	65
Figure 17 Paramètre SMTP SSRS.....	65
Figure 18 Options de remise de rapport SSRS.....	66
Figure 19 Option de paramètres de rapport SSRS.....	66
Figure 20Utilitaire d'exécution de package .....	68
Figure 21 Option de mise en cache d'un rapport .....	72
Figure 22Option de création d'instantanés de rapport .....	72
Figure 23 Flux de contrôle Intégration contrats .....	76
Figure 24 Tâche de flux de données Intégration Contrats .....	76
Figure 25 Modélisation en étoile Fait Séjours.....	79
Figure 26 Schéma étoile séjours .....	80
Figure 27 ETL Insertion DWH SEJOURS .....	81
Figure 28tâche ETL Update Fact Séjour.....	82
Figure 29Chainage exécution tâches ETL .....	82
Figure 30 Fonctionnalités Power Query .....	84
Figure 31Microsoft Composant Power BI .....	85

## Liste des tableaux

Tableau 1 Comparatif caractéristiques types tables de fait .....	20
Tableau 2 Sept types de dimensions à évolution lente .....	22
Tableau 3 Historique marché décisionnel .....	32
Tableau 4 liste des fonctionnalités attendues .....	38
Tableau 5 Liste des formats de données gérés par l'association.....	39

## **Comment une solution décisionnelle peut-elle permettre aux gestionnaires de disposer de données provenant de sources métiers hétérogènes ?**

**Mémoire d'Ingénieur C.N.A.M., Toulouse 2016**

---

### **RESUME**

Le but de ce mémoire est de présenter les concepts de l'informatique décisionnelle permettant la création de rapports et de tableaux de bords. Après y avoir présenté l'association EDENIS et son système d'information, j'y traite de la mise en œuvre des composants décisionnel afin de répondre aux problématiques métiers ainsi que des problématiques d'exploitation.

Puis j'y décris la mise en œuvre des composants sur des domaines pris en exemple

**Mots clés : Système d'information décisionnelle, ETL, entrepôt de données, ODS, SSIS ,SSRS EDENIS.**

---

### **SUMMARY**

The purpose of this thesis is to present the concepts of business intelligence allowing the creation of reports and dashboards.

After having presented EDENIS association and its information system , I dealt with the implementation of the business intelligence component to meet business problems as well as operational issues.

Then I describe the implementation of the components on areas taken as a sample.

**Key words: Decision support system, ETL, Data warehouse, ODS, SSIS,SSRS,EDENIS.**

