



Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Direction Générale des Etudes Technologiques

Institut Supérieur des Etudes Technologiques de Radès  
Département Technologies de l'Informatique

# Rapport de PROJET DE FIN D'ETUDES

Présenté en vue de l'obtention du diplôme de  
Master Professionnel en Business Intelligence

Parcours : MPBI  
Insertion & Employabilité

Elaboré par :  
Fedia Ghobghob Hatira

Encadré par :  
Mme Nesrine Elleuch

Effectué à :  
Entreprise : ISET Radès  
Encadrant : Mme Nesrine Elleuch

Période : Fev - Juin 2022

Année Universitaire : 2021/2022

*Du plus profond de mon cœur, je dédie ce travail à tous ceux qui me sont  
chers,*

## *A mes chers parents*

*Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma  
considération pour les sacrifices que vous avez consentis pour mon instruction  
et mon bien-être.*

*Je vous remercie pour tout le soutien et l'amour que vous me portez depuis  
mon enfance et j'espère que vos bénédictions m'accompagneront pour toujours.  
Que ce modeste travail soit l'exaucement de vos vœux tant formulés, le fruit de  
vos innombrables sacrifices. Puisse Dieu l'omniprésent, l'omnipuissant et  
l'omniscient, vous accorder santé, bonheur et longévité.*

*A mes très chers frères et soeurs ainsi qu'à toute la famille. A mes amis et à  
tous mes proches pour leurs encouragements, leurs assistances et les soutiens  
inconditionnels. A tous les professeurs de l'ISET de Radès, plus  
particulièrement à ceux du département Informatique*



Fedia Ghobghob Hatira

# REMERCIEMENTS

*Il m'est offert ici par quelques lignes la possibilité de remercier les personnes qui ont contribué à faire de ce projet une réussite au sein de l'Institut Supérieur des Etudes Technologique de Radès, un bon stage.*

*Je tiens à remercier et à exprimer ma profonde gratitude à notre chère professeure et encadrante Mme Nesrine Elleuch pour son suivi, accompagnement, son énorme soutien et ses conseils qu'elle m'a prodigués tout au long de ce stage.*

*L'occasion m'est donnée ici de remercier aussi mes camarades de promotion universitaire pour leur esprit de cohésion et de partage qui ont contribué à rendre brillante et enrichissante cette année universitaire.*

# Sommaire

<b>1</b>	<b>Contexte du projet</b>	<b>1</b>
1.1	Introduction . . . . .	1
1.1.1	Présentation de l'organisme d'accueil . . . . .	1
1.1.1.1	Création . . . . .	1
1.1.1.2	Mission . . . . .	1
1.1.1.3	Historique . . . . .	2
1.1.1.4	Ouverture sur l'environnement et convention . . . . .	3
1.1.1.5	Les services de l'ISSET. . . . .	4
1.1.1.6	Principaux Services en relation avec les étudiants . . . . .	5
1.1.2	Cadre du projet . . . . .	5
1.1.3	Description du contexte du projet . . . . .	6
1.1.3.1	Problématique . . . . .	6
1.1.3.2	Étude de l'existant . . . . .	6
1.1.3.3	Identification de l'environnement . . . . .	6
1.1.3.4	Solution proposée . . . . .	7
1.1.4	Contrainte du projet . . . . .	9
1.1.5	Cycle de vie d'un projet décisionnel . . . . .	9
1.1.5.1	Planification . . . . .	9
1.1.5.2	La définition des besoins de l'entreprise . . . . .	9
1.1.5.3	Conception du modèle physique . . . . .	10
1.1.5.4	Modélisation multidimensionnelle . . . . .	10
1.1.5.5	Conception et développement des éléments de la zone de prépa- ration des données . . . . .	10
1.1.5.6	Définition de l'architecture technique . . . . .	10
1.1.5.7	Sélection et installation des outils . . . . .	11
1.1.5.8	Conception de l'application BI . . . . .	11
1.1.5.9	Développement de l'application utilisateur . . . . .	11
1.1.5.10	Déploiement . . . . .	11

1.1.5.11	Maintenance et croissance . . . . .	11
1.1.5.12	Gestion du projet . . . . .	11
1.1.6	Planification du projet . . . . .	11
1.2	Conclusion . . . . .	12
<b>2</b>	<b>Notion Théorique</b>	<b>13</b>
2.1	Introduction . . . . .	13
2.2	Les systèmes décisionnels (Business Intelligence) . . . . .	13
2.2.1	Les avantages d'une solution BI . . . . .	15
2.2.2	Les domaines d'utilisation de la BI . . . . .	15
2.3	Décisionnel vs transactionnel (OLAP vs OLTP) . . . . .	16
2.4	Data Warehouse . . . . .	17
2.4.1	Qu'est Ce qu'un Data Warehouse . . . . .	17
2.4.2	Historique des Data Warehouse . . . . .	17
2.4.3	Modélisation de données d'un DW . . . . .	18
2.4.3.1	Concepts . . . . .	18
2.4.3.2	Différents modèles de la modélisation dimensionnelle . . . . .	18
2.4.4	Concept OLAP . . . . .	20
2.4.5	Architecture et composants d'un système BI . . . . .	21
2.4.5.1	Sources de données . . . . .	21
2.4.5.2	ETL . . . . .	22
2.4.5.3	Entrepôt de données . . . . .	23
2.4.5.4	Modèle de données . . . . .	23
2.4.5.5	Visualisation de données . . . . .	23
2.5	Solutions BI présentes sur le marché . . . . .	24
2.5.1	Environnement logiciel choisi . . . . .	24
2.6	Démarches et Méthodologies . . . . .	26
2.6.1	Méthodologie de travail . . . . .	27
2.6.2	Approches de modélisation BI Adopté . . . . .	27
2.6.2.1	Inmon : de l'entrepôt aux magasins . . . . .	27
2.6.2.2	Kimball : des magasins à l'entrepôt . . . . .	27
2.7	Conclusion . . . . .	28
<b>3</b>	<b>Conception et Développement de l'ETL</b>	<b>29</b>
3.1	Introduction . . . . .	29
3.2	Couche source de données . . . . .	29
3.3	Couche de préparation de données (StagingDataSource) . . . . .	30
3.3.0.1	Extract : Extraction de données brutes . . . . .	30

3.3.0.2	Tansform : Transformation, homogénéisation et nettoyage de ces données . . . . .	35
3.4	Couche de stockage de données (DWH) . . . . .	37
3.5	Couche de présentation et restitution de données . . . . .	38
<b>Bibliographie et Netographie</b>		<b>39</b>
<b>A Annexes</b>		<b>40</b>
<b>Annexes</b>		<b>40</b>

# Table des figures

1.1	Cycle de vie d'un projet decisionnel selon RALPH KIMBALL . . . . .	10
1.2	Diagramme de Gant . . . . .	12
2.1	Le passage de l'informatique de production à la l'informatique décisionnelle . . .	14
2.2	Evolution des bases de données décisionnelles . . . . .	18
2.3	Architecture en étoile qui illustre le concept de fait et le concept de dimension .	19
2.4	Tableau comparatif entre les tables de faits et les tables de dimensions . . . . .	19
2.5	Modélisation en étoile . . . . .	20
2.6	Modélisation en flocons . . . . .	20
2.7	Modélisation en constellation . . . . .	21
2.8	Tableau comparatif des différentes architectures OLAP . . . . .	22
2.9	vue d'ensemble de l'architecture d'un systeme BI . . . . .	22
2.10	SQL Server logo . . . . .	25
2.11	Visual Studio logo . . . . .	25
2.12	Power BI Desktop logo . . . . .	25
2.13	L'approche top-down . . . . .	28
2.14	Approche bottomup . . . . .	28
3.1	Modèle Relationnel de la Base de données Partie relation entre les tables . . . .	30
3.2	Préparation au Staging . . . . .	31
3.3	affichage de l'historique des opérations . . . . .	31
3.4	création des fonctions . . . . .	32
3.5	création dbo.fnFormatMatricule . . . . .	32
3.6	Chargement de la table ImportEtudiant . . . . .	33
3.7	Chargement de la table ImportResultat . . . . .	33
3.8	Modèle Relationnel de la Base de données Partie ImportFichier . . . . .	34
3.9	Transformation et nettoyages de données pour la création de la table Etudiants .	35
3.10	Transformation et nettoyages de données pour la création de la table Resultat .	35
3.11	Transformation et nettoyages de données pour la création de la table Inscription	36

3.12	Modèle Relationnel de la Base de données Partie Etudiant . . . . .	36
3.13	Modèle Relationnel de la Base de données Partie Convention . . . . .	37
3.14	Modèle en étoile (DWH) . . . . .	38



# Liste des tableaux

- 1.1 Tableau récapitulatif des différents rapports requis . . . . . 8
- 2.1 Les différences entre OLTP et OLAP . . . . . 16

# Chapitre 1

## Contexte du projet

### 1.1 Introduction

L'étude préalable constitue une étape primordiale dans le processus de la réalisation d'un projet.

Dans ce chapitre, nous présentons l'organisme d'accueil à savoir l'Institut Supérieur des Etudes Technologiques de Radès (ISET Radès) au sein duquel s'est déroulé notre projet, Nous présenterons par la suite le cadre général du projet, puis un diagnostic de l'existant pour dégager les insuffisances et proposer les orientations de notre future solution.

Nous expliquerons, enfin, le planning de notre projet

#### 1.1.1 Présentation de l'organisme d'accueil

##### 1.1.1.1 Création

L'ISET Radès (Institut supérieur des études technologiques de Radès) créé par la loi 51-92 du 18 mai 1992; est un institut universitaire tunisien rattaché à la direction générale des études technologiques. Placé sous la tutelle du ministère de l'Enseignement supérieur et de la Recherche scientifique et géré par la direction générale, il propose une formation initiale diplômante, des formations continues en lien avec les besoins des entreprises, un centre de ressources technologiques, un pôle de compétence et une pépinière d'entreprises. L'enseignement y est assuré en grande partie par des enseignants du corps des technologues.

##### 1.1.1.2 Mission

ISET-Radès est une institution d'enseignement Supérieur tunisienne formant des étudiants en vue de l'obtention du diplôme de licence appliquée et de master professionnel dans plusieurs domaines. Il assure la formation selon la réforme LMD du système de l'enseignement supérieur

tunisien au sein des départements suivants : génie mécanique, génie civil, génie électrique, technologies de l'informatique et des sciences économiques et de gestion.

La mission de l'ISET de Radès est de « fournir au marché de l'emploi des cadres moyens qualifiés aussi bien dans les secteurs secondaires que tertiaire, de promouvoir le recyclage et la formation continue au profit des cadres exerçant dans les entreprises (formation diplômante en cours de soir, formation spécifique et ponctuelle ) et de mettre en place un partenariat avec les entreprises et les organisations professionnelle à travers la participation des professionnels à la formation et à l'encadrement des étudiants ainsi que la réalisation conjointe de programme de recherche appliquée et de transfert de technologie ».

### 1.1.1.3 Historique

L'Institut Supérieur des Etudes Technologiques de Radès, fleuron du réseau des ISET et porte drapeau de la formation technologique en Tunisie, a constitué depuis sa création en septembre 1995, le principal pourvoyeur du marché de l'emploi en techniciens supérieurs aussi bien dans le secteur industriel que dans le secteur des services. En effet, la localisation de l'ISET de Radès dans l'une des régions à plus forte concentration industrielle du pays, la région de Ben Arous, son intégration dans une zone logistique comprenant le plus grand port tunisien et sa proximité des plus grands pôles de commerce et de services de Tunis ville, ont permis à l'institut de conforter sa vocation régionale et d'assurer une responsabilité grandissante vis-à-vis de ses partenaires professionnels.

Tout au long de ces 30 ans d'existence l'ISET de Radès a été précurseur en termes d'innovation pédagogique et d'ouverture sur l'environnement professionnel. Il a été souvent la scène de prédilection des expériences pilotes touchant le réseau des ISET.

A ce titre l'ISET de Radès a lancé depuis 2003 l'expérience pilote de l'enseignement à distance en Tunisie en collaboration avec l'UVT (départements de Gestion, de techniques de commercialisation et de commerce international), il a été le premier à décrocher un financement dans le cadre du projet d'appui à la qualité de l'enseignement supérieur PAQ (au profit du département de Génie Mécanique, année 2007), et a également mis en œuvre la même années cinq licences co-construites répondant aux besoins immédiats du tissu industriel (une licence co-construite en grande distribution avec la fédération des services, deux licences co-construites en supervision et suivi des chantiers de travaux publics et en supervision et suivi des travaux de bâtiments avec la fédération des entrepreneurs de bâtiments et de travaux publics, et deux autres licences co-construites en essais électriques et électroniques et en systèmes électroniques de sécurité avec la fédération des électriciens FEDELEC).

En outre l'ISET de Radès a décroché le prix de l'innovation technologique par la création

du Centre de Ressources Technologiques en CFAO, mis en place depuis 2005 dans le cadre de la coopération Tuniso-Française (Projet de coopération : Appui au développement de l'enseignement Technologique en Tunisie). Un centre qui se veut un organe de transfert de technologie et d'appui aux entreprises industrielles pour garantir la qualité et les performances indispensables en vue de gagner le pari de la compétitivité.

En septembre 2008 ISET de Radès a intégré le système LMD avec six licences appliquées en Génie Mécanique, en Génie Electrique, en Génie Civil, en Technologies de l'Informatique, en Administration des Affaires et en Marketing débouchant sur 21 parcours permettant la formation de licenciés polyvalents capables de s'adapter aux exigences du marché de l'emploi. Ces licences ont été enrichies en 2014 par une nouvelle filière prometteuse, celle de Finance islamique et en arrivant à ce jour l'ISET atteint 53 parcours.

### 1.1.1.4 Ouverture sur l'environnement et convention

Dans le cadre de sa mission d'ouverture sur l'environnement socio-économique. L'ISET est liée à plusieurs entreprises tunisiennes et internationales par des conventions de partenariat couvrant des contrats de recherche ou des contrats de formation continue. Aussi partenariat avec les entreprises et la société civile pour une meilleure visibilité de l'ISET et des profils des diplômés.

- Participation d'un bon nombre de professionnels issues des milieux industriel et économique à l'enseignement des étudiants de l'**ISET de Radès** ;
- Accueil des étudiants de l'ISET de Radès au sein des entreprises en stages et en projets de fin d'études :
  - 2000 à 2400 stages au sein des entreprises sont affectés chaque année depuis 1999.
  - 400 à 500 projets de fin d'études à caractère professionnel sont aussi proposés chaque année par les industriels pour les étudiants du niveau terminal de l'**ISET Radès**, et encadrés conjointement par un enseignant de l'ISET et un responsable de l'entreprise.
- Participation des industriels aux différentes manifestations et expositions scientifiques et technologiques organisées par l'**ISET de Radès** ;
- Participation de personnalités professionnelles au conseil scientifique et technologique et au comité de direction de l'**ISET Radès** ;
- Participation à la définition des programmes des enseignements ;

La maturation progressive de l'institution associée à l'expérience et la compétence de ses enseignants permettant d'assurer une formation rigoureuse, l'importance des moyens et ressources

technologiques d'une institution en pleine mutation qualitative et l'éventualité de création de pôles de compétences à travers une vague de certification des enseignants permanents et des étudiants. Par ailleurs, sa stratégie s'articule sur le processus de certification de l'école en mettant en œuvre l'SMQ aux termes de l'ISO 9001 en vue de l'accréditation de l'institut prévue à la lumière de l'année universitaire 2021/2022.

En l'occurrence, les étudiants en particulier bénéficient d'un panier de certification (TOEIC, CISCO, C2i, Solid Works, LabView ...) leurs permettant une intégration meilleure dans le secteur professionnel. En outre, l'institut est bien équipé pour répondre d'une manière satisfaisante aux exigences de la formation pratique (travaux pratiques) dans les différents parcours déjà habilités. Les éléments sus indiqués constituent des signes prometteurs en vue d'une revalorisation du statut de notre institut pour une réponse plus pointue aux besoins des professionnels aussi bien en cadres moyens (licence appliquées) qu'en cadres supérieurs (masters professionnels).

### **1.1.1.5 Les services de l'ISET.**

L'organisation de l'ISET est présentée sous forme de l'administration qui capitalise plusieurs services Notamment :

- Services des études et de Stage.
- Secrétariat Général.
- Conseil Scientifique.
- Services Scolarité.
- Service Examen.
- Service Financier et Enseignant.
- Service de Resource Humain.
- Bibliothèque.
- Service de maintenance.
- Les départements.
- Bureau d'Ordre.
- Service partenarial, relations avec l'environnement et formation continue.
- Service de Tirage.
- L'infirmierie.
- Bureau de conseil, d'assistance et d'intervention psychologique.

### 1.1.1.6 Principaux Services en relation avec les étudiants

- **Service Scolarité** : qui a pour rôle la préparation de l'inscription notamment :
  1. faire l'inscription des nouveaux (nouveau bachelier) et anciens étudiants et les affecter dans des groupes de classe ;
  2. fournir les attestations des inscriptions, les attestations de présence, les attestations de départ, les attestations de retrait et toute autres attestations demandé par l'étudiant ;
  3. Préparation et remise des diplômes ;
- **Service de stages** : a pour rôle de faire le suivi des notes de stages d'initiation, perfectionnement ou de PFE, les partager sur la plateforme **DIDOSOFT** pour qu'elle soit visible pour les étudiants et le service Examen.
- **Service Examen** : qui a pour rôle la préparation des examens et suivi des résultats notamment :
  1. recueille les notes des étudiants de différents départements à travers le système DIDOSOFT ou manuellement ;
  2. saisir les notes, calcul les moyennes et faire sortir les PVs ;
  3. fournir les relevés et les attestations de réussites ;
- **Départements** qui jouent un rôle pédagogique avec les étudiants.
- **Bureau de conseil, d'assistance et d'intervention psychologique.**
- **l'infirmierie**

### 1.1.2 Cadre du projet

Notre projet s'inscrit dans le cadre d'un projet de fin d'études pour l'obtention du Mastère Professionnel en Business Intelligence. Le but principal de l'institut est de garder un historique complet sur leurs étudiants même faire un suivie toute au long de leurs cursus scolaires dans le but de connaître quelles sont les parcours et les spécialités qui ont un bon pourcentage de réussite, Aussi un suivie concernant les conventions fournit à l'ISET. Et ce afin d'y remédier et assurer l'amélioration du contenue de la formation.

**Motivation** En étant qu'étudiante en master Business Intelligence, à l'ISET Rades, j'ai voulu contribuer à l'évolution de notre établissement en créant des tableaux de bord permettant de piloter certaines des activités de l'institut. En effet :

L'ISET souffre du manque d'outils décisionnels qui fournissent les différents éléments d'information nécessaires pour avoir une vision historisée des différentes données et permettre d'éventuelles prises de décision.

### 1.1.3 Description du contexte du projet

Cette section a pour objectif de présenter le contexte du projet, de poser la problématique et d'étudier l'existant afin d'exposer ma solution.

#### 1.1.3.1 Problématique

La décision nécessite une matière première : la donnée ou l'information. Si l'information est incomplète, floue, imprécise..., probablement, les décisions seront approximatives, inadaptées, voire fatales ; par contre si l'information est exacte, complète et pertinente, les décisions ont davantage de chances d'avoir une efficacité garantie. Justement, le manque de cette matière première sous forme d'information utile ou bien la difficulté de la faire ressortir, constitue aujourd'hui, une vraie problématique pour les responsables au sein de l'ISSET pour l'élaboration des rapports de gestion, l'édition des graphes et statistiques et les problèmes de prise de décisions en général.

#### 1.1.3.2 Étude de l'existant

Dans cette section nous présentons notre projet en commençant par une étude de l'existant en identifiant l'environnement logiciel dans lequel notre étude va être menée et analyser les structures des données du logiciels de gestion de scolarité et de note de l'ISSET, ensuite nous allons enchaîner par la solution proposée et la méthodologie adoptée pour réussir notre mission.

#### 1.1.3.3 Identification de l'environnement

En analysant la situation actuelle, on a pu distinguer deux systèmes sources qui sont : **SALIMA** (Système d'Administration des LIcences et MAstères), et un autre logiciel nommé 'GestionNote' & 'Scolarité' spécialement pour la gestion d'inscription et gestion de note du parcours mastères. Ce logiciel est élaboré par un enseignant universitaire. "SALIMA" comporte trois axes.

- Le premier est relatif à la gestion pédagogique des filières scientifiques dans les divers établissements universitaires en ce qui concerne la définition des contenus de la formation et le suivi de leur application.
- Le second concerne la gestion des services universitaires et des affaires estudiantines telles que l'inscription universitaire, le système de notation, la carte d'étudiant, la carte de bibliothèque, le relevé des notes ainsi que les attestations de réussite et les diplômes.
- Le troisième axe porte sur la gestion des examens universitaires.

Le ministre a précisé que l'application "SALIMA" intervient en vue de faciliter la gestion des établissements universitaires conformément au régime **LMDbf** (Licence Mastère Doctorat).

Mais comme tout système peut avoir des avantages et des inconvénients, **SALIMA** et l'autre logiciel ont les lacunes suivantes :

- Pas de dictionnaire de données.
- Pas de tableau de bord.
- Format des fichiers incohérent.
- Pas de traçabilité des données.

### 1.1.3.4 Solution proposée

C'est pour ça nous proposons la mise en place d'une solution BI qui couvre l'ensemble de processus allant de l'historisation des données jusqu'à la diffusion d'une vue globale sur un tableau de bord clair permettant ainsi une analyse pertinente et précise.

La démarche de cette solution est planifiée en 3 étapes :

- La première étape consiste à rassembler les données à partir du système SALIMA et le système du parcours mastère concernant les étudiants encore présents, à les traiter et à les stocker, l'objectif est de garder un historique complet des étudiants. Ainsi que le rassemblement des données concernant les conventions de l'ISET.
- La deuxième volet porte sur la production des rapports destinés aux décideurs de l'institut pour avoir une vision détaillée et explicite sur les étudiants en question.

Dans les tableaux de bord, nous devons présenter des indicateurs qui permettront de suivre :

1. Les étudiants au sein de l'ISET dès leurs inscriptions jusqu'à l'obtention du diplôme.
2. Les conventions, leurs durées, les thèmes et les partenaires qui ont signé plus grand nombre convention avec l'ISET.

Ces indicateurs sont classés dans le tableau suivant qui nous a permis de recenser les différents rapports dont les décideurs ont besoin :



Rapport	Catégorie	Description
Nombre des étudiants qui ont abandonnés	Histogramme	Cet histogramme représente le nombre des étudiants qui ont abandonné leur parcours sur l'axe y et les différents parcours sur l'axe x
Nombre des étudiants diplômés par département	Histogramme	Cet histogramme montre le nombre des étudiants qui ont obtenu leurs diplômes, soit dans la série licence ou master : l'axe y représente le nombre des étudiants, l'axe x représente le type du diplôme obtenu par l'étudiant (licence, master).
Nombre des étudiants par département	Histogramme	Cet histogramme permet aux décideurs de visualiser les effectifs des étudiants par département, sur l'axe y on trouve le nombre des étudiants, sur l'axe x les différents départements.
Nombre des étudiants par observation	Histogramme	Cet histogramme montre le nombre d'étudiants selon leur observation (Admis ou Redouble)
Nombre des étudiants par Spécialité	Histogramme	Cet histogramme permet de visualiser les effectifs des étudiants par spécialité, sur l'axe y on trouve le nombre des étudiants et sur l'axe x les différentes spécialités.
Effectif des étudiants par type diplôme	Camembert	Ce camembert affiche l'effectif des étudiants selon leurs diplômes (licence, master).
Effectif des étudiants par type département	Camembert	Ce camembert affiche l'effectif des étudiants par département
Nombre des étudiants par gouvernorat	Carte	Ce rapport sous forme d'une carte géographique expose le nombre des étudiants selon leurs gouvernorat d'origine.
Nombre des étudiants par genre	Table	Cette table affiche le nombre des étudiants selon leur genre (Homme ou Femme)
Nombre des étudiants par état	Table	Cette table affiche le nombre des étudiants selon leurs état (Confirmé, Retrait, Départ)
Taux de Réussite par département	Table	Cette table affiche le pourcentage de réussite des étudiants par département
Taux de Réussite par spécialité	Table	Cette table affiche le pourcentage de réussite des étudiants par spécialité
Nombre de convention par type	Histogramme	Cet histogramme représente le nombre des conventions par type
Nombre de convention par thème	Histogramme	Cet histogramme représente le nombre des conventions par thème
Durée moyenne des conventions	Table	Cette table affiche la durée moyenne des conventions par type ou thème
Nombre de partenaires	Table	Cette table affiche le nombre de partenaire qui ont signé une convention

TABLE 1.1 – Tableau récapitulatif des différents rapports requis

### 1.1.4 Contrainte du projet

Un projet décisionnel peut subir des difficultés il est nécessaire d'évaluer les facteurs qui peuvent ralentir le déroulement du projet. Après l'étude de l'existant on s'est opposé à une multitude de problèmes et difficultés majeures qui nous ont retardés dans notre démarche.

La liste suivante contient une liste non exhaustive des obstacles et des problèmes les plus importants :

- Manque d'historisation des données ;
- Manque de Bases de données ;
- Données non structurées ;
- Tables non référencées (pas de relations entre les tables) ;
- Champs de tables non renseignés (vide) ou partiellement renseignés ;
- Champs remplis de manière erronée ;
- Manque de données pour faire exécuter les requêtes.
- Champs redondants (Même champs retrouvé plusieurs fois dans la base de données)

### 1.1.5 Cycle de vie d'un projet décisionnel

L'implémentation d'un système Business Intelligence/Intelligence d'affaires doit être pilotée par une méthodologie globale. Une méthodologie qui diffère des méthodologies de développement des systèmes transactionnels.

Pour garantir l'aspect fonctionnel du projet et avoir des résultats rapides nous suivrons une méthodologie par cycles itératifs de 'Ralph Kimball'. Cette méthode est illustrée par le schéma suivant :

Pour en savoir plus pour chaque phase et chaque étape du projet je vais présenter les différentes phases dans l'ordre :

#### 1.1.5.1 Planification

La planification aborde la définition et l'étendue du projet de l'entrepôt, elle se focalise sur les besoins en termes de ressources et de niveau de qualification, couplées aux affectations des tâches, à leurs durées et à leur séquençement.

#### 1.1.5.2 La définition des besoins de l'entreprise

Comprendre les attentes du client et de ses exigences non énoncées, poser les bonnes questions et tirer l'information du client. Ce niveau va préparer le terrain du travail en termes de 3

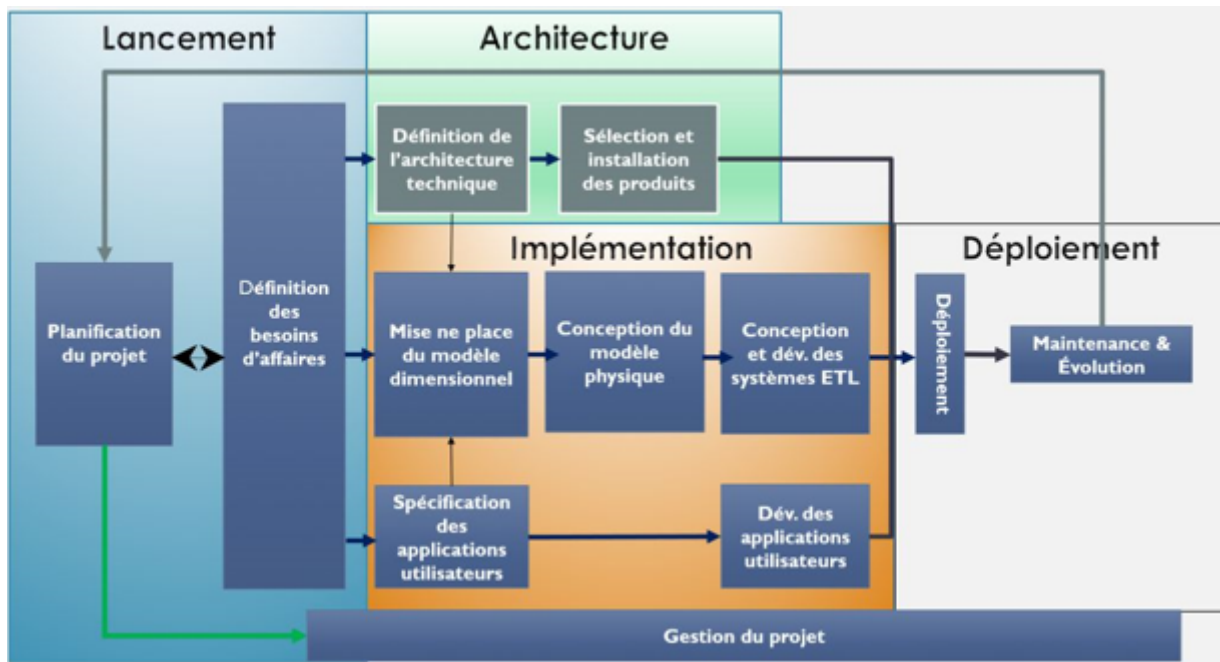


FIGURE 1.1 – Cycle de vie d'un projet décisionnel selon RALPH KIMBALL

trajectoires ; La technologie, le modèle et l'interface client.

### 1.1.5.3 Conception du modèle physique

La conception physique d'une base de données définit les structures nécessaires pour l'implémentation du modèle dimensionnel.

### 1.1.5.4 Modélisation multidimensionnelle

Elaboration du modèle multidimensionnel des données : faits et dimensions et la réalisation des maquettes de la solution décisionnelle proposée.

### 1.1.5.5 Conception et développement des éléments de la zone de préparation des données

Concevoir et développer les étapes d'ETL qui vont résulter des données nettoyées et atomiques alimentant notre magasin de données.

### 1.1.5.6 Définition de l'architecture technique

Cette étape définit la vision globale de l'architecture technique à mettre en oeuvre. Elle nécessite la prise en compte de trois facteurs : Les besoins ; L'environnement existant et les orientations techniques stratégiques planifiées.

### **1.1.5.7 Sélection et installation des outils**

Réaliser une étude comparative des outils pour chaque phase de la chaîne décisionnelle.

### **1.1.5.8 Conception de l'application BI**

Il s'agit de définir une série d'applications standard destinées aux utilisateurs finaux. Les spécifications de l'application décrivent les maquettes d'états, les critères de sélection laissés à l'utilisateur et les calculs nécessaires.

### **1.1.5.9 Développement de l'application utilisateur**

Développer les tableaux de bord et indicateurs de performance spécifique pour chaque utilisateur.

### **1.1.5.10 Déploiement**

Intégrer après validation et test, l'application au sein de l'organisme.

### **1.1.5.11 Maintenance et croissance**

Il faut s'assurer de fournir un service de support et de formation continue. Il est également important de mesurer périodiquement l'évolution et la croissance des performances de l'entrepôt de données au sein de l'entreprise.

### **1.1.5.12 Gestion du projet**

Garantir le bon déroulement des activités du cycle de vie dimensionnel, contrôler l'état d'avancement du projet et enfin détecter et résoudre des problèmes. Dans l'obligation de trouver une méthode qui englobe toutes les phases du projet permettant de satisfaire le client dans les plus brefs délais, nous allons utiliser la méthodologie « Ralph KIMBALL » pour notre projet vue qu'elle s'adapte au mieux à nos besoins en termes de compétences, de temps, de coût et d'exigence en ce qui concerne l'intégration de données.

## **1.1.6 Planification du projet**

Le développement de tout projet nécessite une phase de planification qui permet de définir les tâches à réaliser, maîtriser les risques et rendre compte de l'état d'avancement du projet. Dans le but d'une meilleure organisation du projet et gestion efficace des délais disponibles, j'ai présenté sous forme de diagramme de GANTT qui illustre cet ordonnancement.

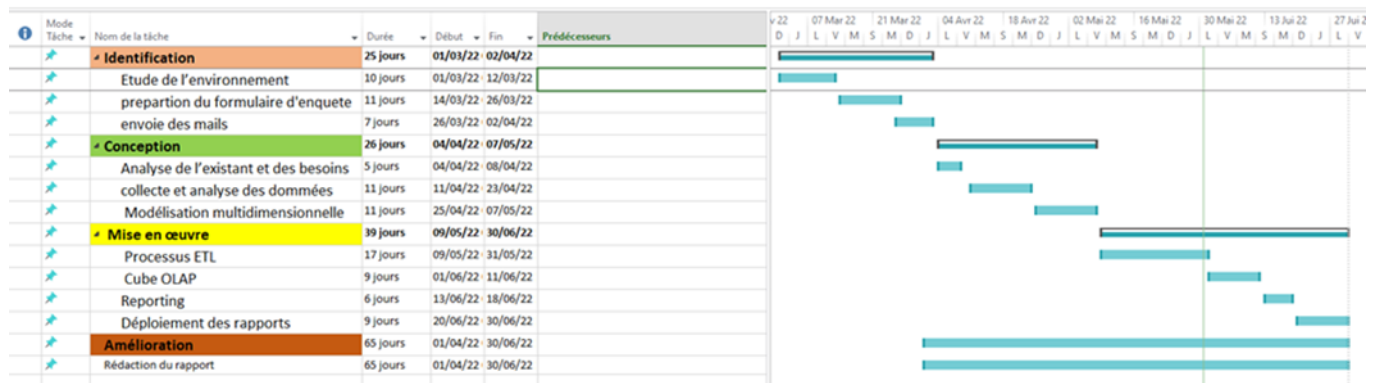


FIGURE 1.2 – Diagramme de Gant

## 1.2 Conclusion

Ce premier chapitre a fait l'objet d'une présentation de l'existant, le cadrage du projet par l'élaboration de sa problématique et les différents objectifs.

A la fin de ce chapitre j'ai spécifié un cycle de vie décisionnel standardisé assurant la réalisation du projet ainsi que la planification.

# Chapitre 2

## Notion Théorique

### 2.1 Introduction

Toutes les entreprises du monde possèdent plus ou moins d'énormes quantités de données. Ces informations proviennent de sources internes (générées par le système d'exploitation dans ses activités quotidiennes) ou de sources externes (web, fichiers plats, etc.)

Cette surabondance des données d'une part, et les limites des systèmes opérationnels pour exploiter cette quantité de données à des fins d'analyse d'une autre part, ont conduit et pousser les entreprises à tourner vers une nouvelle ère informatique dite décisionnelle (Business Intelligence).

L'informatique décisionnelle, également Business Intelligence ou BI en anglais, désigne les moyens, les méthodes et les outils qui apportent des solutions en vue d'offrir une aide à la décision aux professionnels afin de leurs permettre d'avoir une vue d'ensemble sur l'activité de l'entreprise et de leurs permettre de prendre des décisions plus avisées à travers des tableaux de bord de suivi et des analyses.

### 2.2 Les systèmes décisionnels (Business Intelligence)

La Business Intelligence désigne les moyens, les outils et les méthodes qui permettent de collecter, consolider, modéliser et restituer les données d'une entreprise en vue de fournir une aide à la décision aux managers.

Le terme français est « Informatique Décisionnelle ». Une application de ce genre exécute la capture, l'analyse et le stockage de données provenant de plusieurs sources hétérogènes qui peuvent être des Enterprise Resource Planning (ERP), des bases de données ou d'autres entrepôts de

données. Traditionnellement, un entrepôt de données est utilisé comme source d'information par les décideurs. La Business Intelligence s'insère dans l'architecture du système d'information d'une entreprise.

Avec Business Intelligence, nous comprenons les applications Business Intelligence. Cependant, le terme « business intelligence » ne fait pas directement référence à des outils informatiques, mais à ce que l'on fait avec ces outils afin d'effectuer des analyses poussées sur des objets préalablement identifiés. Cette analyse peut être établie pour l'ensemble de l'entreprise, pour un département ou une division spécifique, ou même pour un projet spécifique.

Pour pouvoir obtenir une vue complète des objets analytiques de haut niveau, les applications d'informatique décisionnelle utilisent des entrepôts de données comme sources d'informations. La tâche principale d'un entrepôt de données est de filtrer, recouper et recatégoriser les informations pour fournir un générateur d'analyse de données au niveau mondial. C'est-à-dire des données pertinentes directement liées à l'objet analysé. Dans la plupart des cas, ces données proviennent de bases de données relationnelles. Un entrepôt de données est un clone de données existantes au niveau opérationnel d'une entreprise en préparation de problèmes d'analyse et de performance. Les images des situations de l'entreprise ou les résultats produits par les générateurs d'analyses sont présentés sous forme de rapports statistiques ou sous forme de tableaux de bord décisionnels. Ces informations sont utilisées pour vérifier l'alignement de la stratégie de l'entreprise et assister la direction de l'organisme, d'où le nom d'informatique décisionnelle.

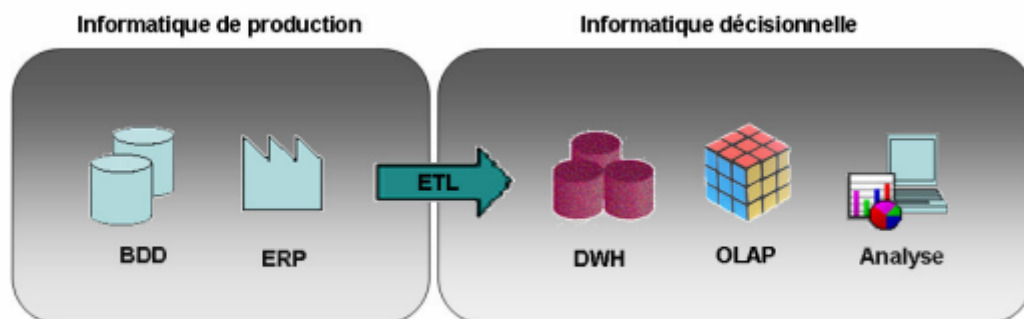


FIGURE 2.1 – Le passage de l'informatique de production à la l'informatique décisionnelle

Le passage de l'informatique de production à l'informatique décisionnelle se fait par un Extract Transform Load (ETL). Les sources de données les plus utilisées dans ce genre de transaction sont les bases de données. Les informations contenues dans la source de données sont traitées puis stockées dans un entrepôt de données. Les données gardées dans l'entrepôt de données sont dans l'état de consultation, et sont sujettes à l'analyse multidimensionnelle (OnLine Analytical Processing OLAP). Le résultat produit par l'analyse multidimensionnelle est appelé « cube ». Les cubes d'information sont traités par les générateurs d'analyse pour

produire les reports demandés. [7]

### 2.2.1 Les avantages d'une solution BI

Sans aucun doute, le terme "BI" est devenu l'un des mots à la mode les plus utilisés dans le monde du business. Les nouvelles solutions proposées sont faciles à utiliser avec une interface ergonomique simple. Par contre, la question qu'il faut se poser est : "Est-ce que mon entreprise a vraiment besoin d'une solution BI?". Si la réponse est oui : "Comment cela affecte-t-il mes décisions au sein de l'entreprise?".

Voilà les bénéfices apportés couramment, en implémentant une solution BI au sein d'une organisation :

- Rapidité de la prise de décision fondée sur des faits.
- Améliorer la visibilité sur les chiffres, les écarts et les anomalies.
- La combinaison de plusieurs sources de données (ERP, systèmes comptable, feuilles de calcul, des budgets ...).
- La présentation uniforme d'informations fiables pour les membres de l'organisation.
- L'automatisation permettant l'accélération de la collecte et de la diffusion de l'information.
- La performance dans le calcul d'agrégats sur de gros volume de données.
- La prise de décision grâce à des indicateurs pertinents et à une structure cohérente des informations.
- L'aide à nettoyer les données présentes dans différents logiciels.
- L'anticipation des événements et la projection dans l'avenir.
- Elaboration des rapports nécessaires au bon moment, à tous les niveaux organisationnels, et dans les meilleurs formats possibles.

### 2.2.2 Les domaines d'utilisation de la BI

La **BI** est utilisée dans plusieurs domaines nous citerons :

- Finance, avec le reporting financiers et budgétaires.
- Vente et commercial, avec l'analyse de points de vente, l'analyse de la rentabilité.
- Marketing, avec la segmentation clients, les analyses comportementales.
- Logistique, avec l'optimisation de la gestion de stock, le suivi des livraisons.
- Ressources humaines, avec l'optimisation de l'allocation des ressources.



## 2.3 Décisionnel vs transactionnel (OLAP vs OLTP)

Les systèmes informatiques peuvent se subdiviser en deux catégories : les systèmes transactionnels OLTP (Online Transaction Processing) et les systèmes analytiques OLAP (OnLine Analytical Processing).

- Les systèmes OLTP sont dédiés aux métiers de l'entreprise pour les assister dans leurs tâches de gestion quotidiennes et donc directement opérationnels. Le mode de travail est transactionnel. L'objectif est de pouvoir insérer, modifier et interroger rapidement et en sécurité la base. Ces actions doivent pouvoir être effectuées très rapidement par de nombreux utilisateurs simultanément. Il est proposé essentiellement pour les applications gérant des opérations commerciales comme les opérations bancaires, ou l'achat de bien divers. [4]
- Les systèmes OLAP sont dédiés au management de l'entreprise pour l'aider au pilotage de l'activité. C'est un outil de reporting dont la couche d'analyse permet de générer des résultats en fonction du contenu d'un entrepôt de données. Les programmes consultent une quantité importante de données pour procéder à des analyses. Les objectifs principaux sont : regrouper, organiser des informations provenant de sources diverses, les intégrer et les stocker pour permettre à l'utilisateur de retrouver et analyser l'information facilement et rapidement. [4]

Bien que les systèmes d'informations **OLTP** et **OLAP** aient le point commun de regrouper les données de l'entreprise dans un SGBD et d'en fournir l'accès aux utilisateurs, ils présentent de profondes différences, présentées dans ce tableau :

Caractéristiques	OLTP	OLAP
Utilisation	SGBD base de production	Entrepôt de donnée
Opération Typique	Mise à jour	Analyse
Type d'accès	Lecture écriture	Lecture
Taille BD	Faible (max quelque GB)	Importante (pouvant aller à plusieurs TB)
Requête	Simple, Régulières, Répétitives, Nombreuses	Complexes, Irrégulières, peu nombreuses, non prévisibles
Quantité d'informations échangées	Faible	Important
Orientation	Ligne	Multi dimension
Structure de données	Beaucoup de tables	Peu de table mais de grande tailles
Ancienneté des données	Récente	Historique)

TABLE 2.1 – Les différences entre OLTP et OLAP

Ces nouvelles exigences des entreprises ont motivé la nécessité de mettre en place un système répondant aux besoins décisionnels. Ce système n'est rien d'autre que le «Data Warehouse».

## 2.4 Data Warehouse

### 2.4.1 Qu'est Ce qu'un Data Warehouse

Un Data Warehouse ou un entrepôt de données est une vision centralisée et universelle de toutes les informations de l'entreprise. C'est une structure (comme une base de données) qui a pour but, contrairement aux bases de données, de regrouper les données de l'entreprise pour des fins analytiques et pour aider à la décision stratégique. Tous les types de Data Warehouse partagent les cinq caractéristiques suivantes :

- Données organisées ;
- Données cohérentes ;
- Données évolutives dans le temps (les données sont conservées sur une période plus longue, cela permet la comparaison et le suivi de l'évolution des valeurs dans le temps) ;
- Non volatiles (les données ne peuvent pas être mises à jour, les données sont historisées) ;
- Structure relationnelle. [9]

### 2.4.2 Historique des Data Warehouse

Les origines du concept de "datawarehouse" DW (entrepôt de données en français) remontent aux années 1980, période pendant laquelle il y avait un intérêt croissant pour les systèmes décisionnels, principalement en raison de l'avènement du SGBD relationnel simple et puissant fourni par le langage SQL.

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

Le Data Warehouse, tel qu'on le connaît actuellement, n'est plus vu comme une copie ou un cumul de copies prises de façon périodique des données opérationnel mais une nouvelle source d'information, alimentée avec des données recueillies et consolidées des différentes sources internes et externes. [5] Le stockage en entrepôts des données a commencé à la fin des années 1980 lorsque Paul Murphy et Barry Devlin, employés d'IBM, a mis au point le Business Data Warehouse.

Cependant, le vrai concept a été donné par Inmon Bill. Il était considéré comme le père de

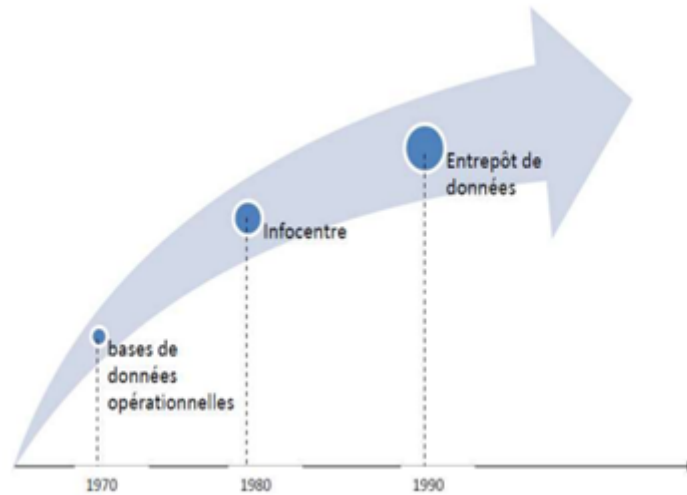


FIGURE 2.2 – Evolution des bases de données décisionnelles

l'entrepôt de données. Il avait écrit sur une variété de sujets concernant la construction, l'utilisation et l'entretien de l'entrepôt et de l'usine d'information corporative.

### 2.4.3 Modélisation de données d'un DW

La modélisation dimensionnelle d'un DW s'appuie sur les deux concepts suivants :

#### 2.4.3.1 Concepts

- **Concept de fait** : Les tables de fait contiennent les données que l'on souhaite voir dans des rapports d'analyse. Une table de faits se présente sous la forme d'un ensemble de colonnes stockant des valeurs dites mesures, et des clés étrangères (identifiants) qui sont généralement les clés primaires associées aux tables de dimensions.
- **Concept de dimension** : Les tables de dimension sont les tables qui raccompagnent une table de faits, elles contiennent les descriptions textuelles de l'activité. Une table de dimension est constituée de nombreuses colonnes qui décrivent une ligne. C'est grâce à cette table que l'entrepôt de données peut être compréhensible et utilisable [8].

Le tableau suivant résume les différences entre les tables de faits et les tables de dimensions au niveau des données [1] :

#### 2.4.3.2 Différents modèles de la modélisation dimensionnelle

Principalement, les trois modèles, couramment employés pour la modélisation dimensionnelle sont : **Modèle en étoile**

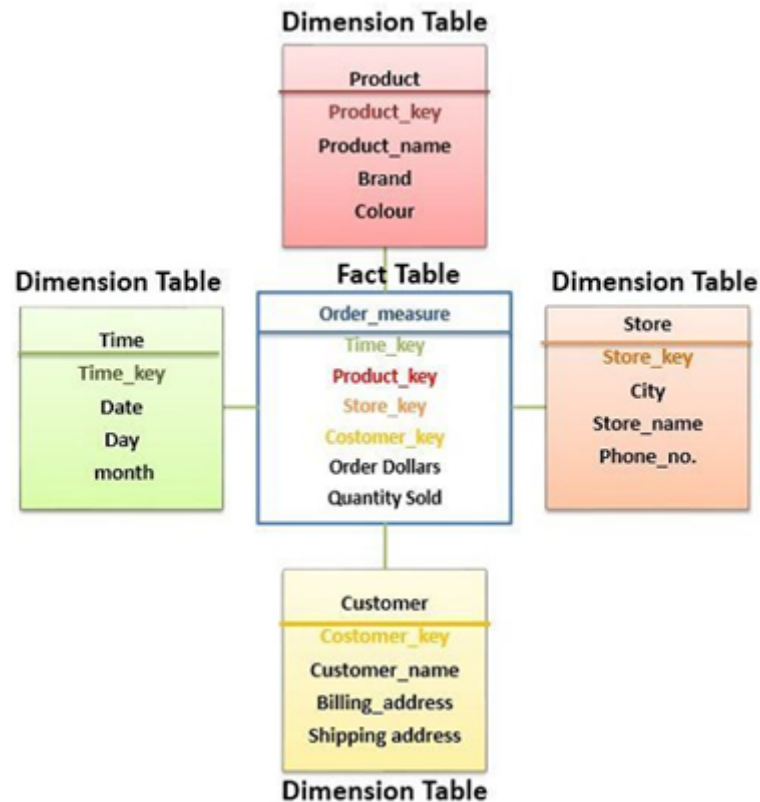


FIGURE 2.3 – Architecture en étoile qui illustre le concept de fait et le concept de dimension

	Tables de faits	Tables de dimensions
<b>Structure</b>	Peu de colonnes beaucoup de lignes	Peut de lignes beaucoup de colonnes
<b>Données</b>	Mesurable, généralement numérique	Descriptives généralement textuelles
<b>Référentiel</b>	Plusieurs clés étrangères	une clé primaire
<b>Valeur</b>	Prend de nombreuses valeurs	Plus ou moins constantes
<b>Manipulation</b>	Participe a des calculs	Participe à des contraintes
<b>Signification</b>	Valeur de mesures	Descriptive
<b>Rôle</b>	Assure les relations entre les dimensions	Assure l'interface homme/entrepôt de données

FIGURE 2.4 – Tableau comparatif entre les tables de faits et les tables de dimensions

Ce modèle tire son nom de sa configuration, en effet il se forme d'un objet central qui est la table des faits reliée à un ensemble de tables de dimensions. La force de ce type de modélisation est sa lisibilité et sa performance. Présenté dans le schéma ci-dessous :

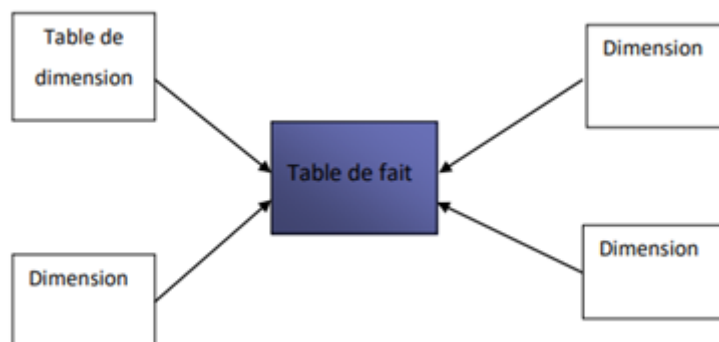


FIGURE 2.5 – Modélisation en étoile

### Modèle en flocon

Identique au modèle en étoile, sauf que ses branches sont éclatées en hiérarchies. Cette modélisation est généralement justifiée par l'économie d'espace de stockage, cependant elle peut s'avérer moins compréhensible pour l'utilisateur final et très coûteuse en termes de performances.

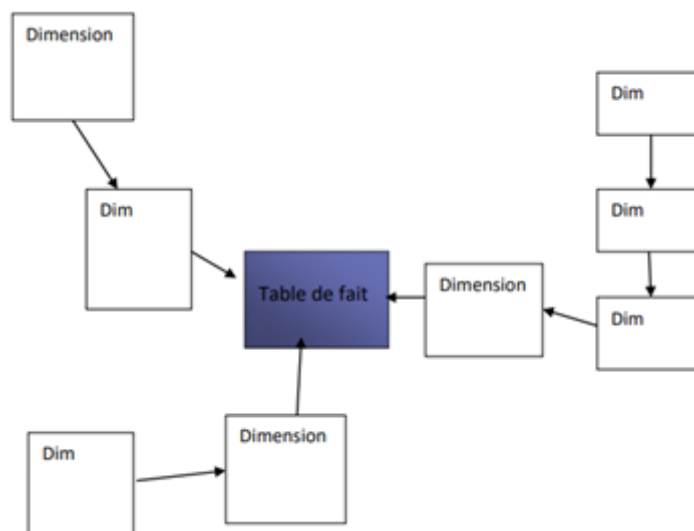


FIGURE 2.6 – Modélisation en flocons

### Modèle en constellation

Ce n'est rien d'autre que plusieurs modèles en étoiles liés entre eux par des dimensions en commun.

## 2.4.4 Concept OLAP

C'est une catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation.

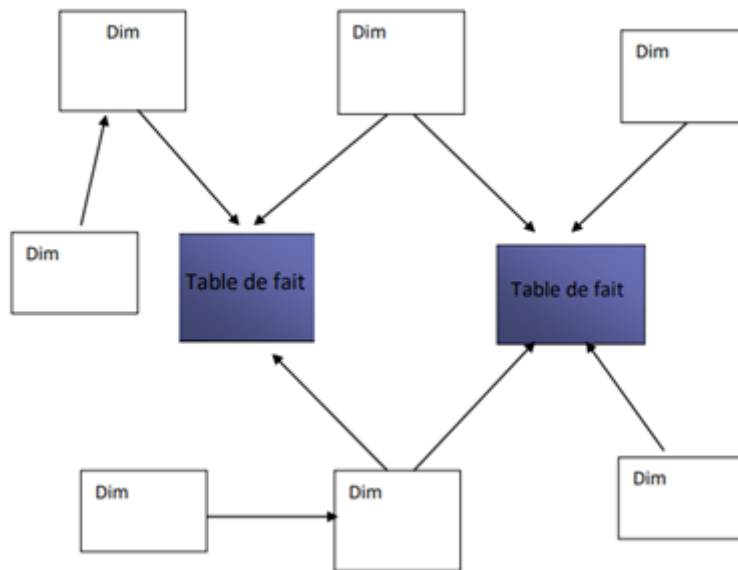


FIGURE 2.7 – Modélisation en constellation

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

- Les systèmes à architecture MOLAP «Multidimensional On-line Analytical Processing»
- Les systèmes à architecture ROLAP «Relationnel On-line Analytical Processing»
- Les systèmes à architecture HOLAP «Hybride On-line Analytical Processing [1]»

## 2.4.5 Architecture et composants d'un système BI

Après avoir défini la BI, on va découvrir ses composants et expliquer comment ça fonctionne les uns avec les autres. Le schéma suivant illustre l'architecture et les composants principaux d'un système BI [1] :

L'architecture présentée dans la figure ci-dessus contient les composants les plus courants trouvés dans la plupart des systèmes de BI, car l'architecture de ses composants peut varier en fonction de l'outil, de l'environnement, etc. Chacun de ces composants seront discutés plus en détail dans les sections qui suivent.

### 2.4.5.1 Sources de données

Les sources de données sont souvent diverses et variées on peut citer par exemple : les bases de données relationnelles (MySQL, Oracle, SQL Server), les fichiers plats Excel, XML, les services web... etc. On ne peut pas construire un système BI, sans source de données.

	ROLAP	MOLAP	HOLAP
Stockage des données de base	BD relationnelle	BD multidimensionnelle (hyper cube)	BD relationnelle
Structure de la BD	Modèle particulier (étoile, flocon, etc)	Similaire à ROLAP la différence est au niveau d'entrepotage des données	Croisement des architectures ROLAP et MOLAP
Fonctionnement	Le serveur extrait les données par des requête SQL et interprète les données selon une vue multidimensionnelle avant de les présenter au module client.	Le serveur MOLAP extrait les données de l'hyper cube et les présente directement au module client.	Accède aux deux BD et les présente au module client selon leur méthode respective
Performance des requêtes	Le moins performant	Le plus performant	Performance moyenne

FIGURE 2.8 – Tableau comparatif des différentes architectures OLAP

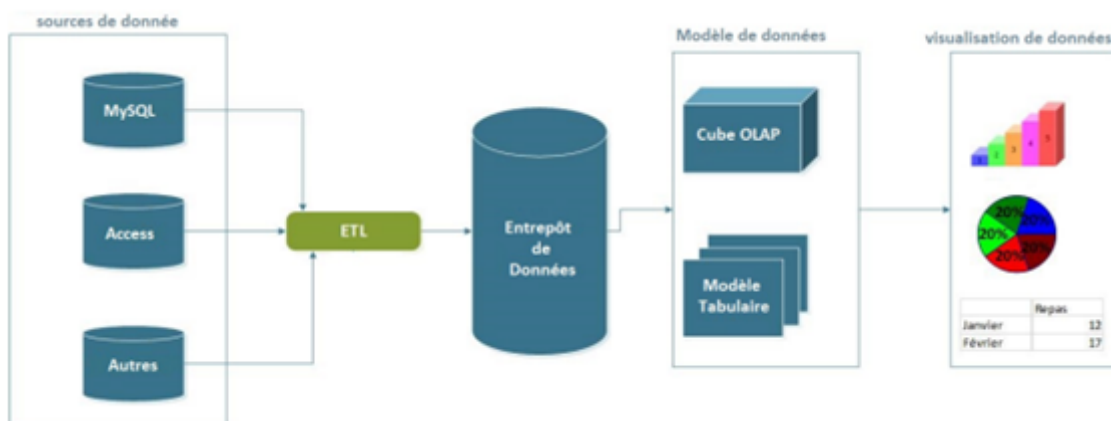


FIGURE 2.9 – vue d'ensemble de l'architecture d'un système BI

#### 2.4.5.2 ETL

Les sources de données sont souvent diverses et variées on peut citer par exemple : les bases de données relationnelles (MySQL, Oracle, SQL Server), les fichiers plats Excel, XML, les services web... etc. On ne peut pas construire un système BI, sans source de données.

### 2.4.5.3 Entrepôt de données

Considéré comme étant le cœur du système vu son importance, il sert à conserver toutes les données de l'entreprise provenant de ses différents services, dans le but de faciliter l'analyse des données et le Reporting.

### 2.4.5.4 Modèle de données

Un Entrepôt de données est conçu pour être la source d'analyse et de rapports, de sorte qu'il soit plus rapide que les systèmes opérationnels pour la production de rapports. Cependant, il n'est pas aussi rapide qu'il prétend, pour pouvoir couvrir tous les besoins, puisque lui-même est une base de données relationnelle qui est définie avec un ensemble de contraintes qui réduisent le temps de réponse d'une requête. L'exigence d'un traitement plus rapide en ayant un temps de réponse plus faible d'une part, et avoir des informations agrégées d'autre part, entraîne donc la création d'une autre couche dans les systèmes de BI. Cette couche, appelé Modèle de données, contient un modèle de données basé sur des fichiers ou basé sur la mémoire vive, afin de produire rapidement les réponses aux requêtes pour l'édition des rapports??.

Le **Modèle de données** s'appuie sur deux technologies : **Cube OLAP** et le modèle en mémoire de tableau (**Modèle tabulaire**).

Le **Cube OLAP** est un stockage de données basé sur un fichier qui charge les données à partir d'un entrepôt de données dans un modèle de cube. Le cube contient des informations descriptives telles que les dimensions (par exemple, Etudiant et Parcours) et des cellules (par exemple, des faits et des mesures, telles que FactResultat et mesure les Admis).

Le Modèle tabulaire est basé sur un nouveau moteur en mémoire pour les tables. Le moteur en mémoire charge toutes les données rangées de tables dans la mémoire et répond à des questions directement à partir de la mémoire. C'est très rapide en termes de temps de réponse.

### 2.4.5.5 Visualisation de données

L'interface d'un système BI?? est la visualisation de données. En d'autres termes, la visualisation de données est la partie du système de BI que les utilisateurs peuvent voir. Il existe différentes méthodes pour la visualisation d'informations, tels que des tableaux de bord stratégiques et tactiques, indicateurs clés de performance ou Key performance Indicators (KPI), des rapports consolidés.

C'est donc la partie qui intéresse le plus les décideurs d'une entreprise. De ce fait, il existe un grand nombre d'outils de visualisation de rapport sur le marché, telle que **SharePoint**, **PowerBI**, **IDashBoards**, etc.



## 2.5 Solutions BI présentes sur le marché

De nombreux fournisseurs de solutions BI tentent de convaincre les clients des performances de leurs applications. Cependant, ces solutions de BI peuvent ne pas convenir à différents environnements commerciaux. Dans la liste ci-dessous, vous pouvez voir les principaux fournisseurs de BI sur le marché. Destinés principalement aux entreprises et aux sociétés, ces produits comprennent une suite complète d'applications, de bases de données, d'outils d'intégration et de visualisation.

- IBM Cognost
- SAP Business Objects
- Pantaho BI
- SpagoBI
- JasperSoft
- SAS BI
- Oracle BI
- Microsoft BI

### 2.5.1 Environnement logiciel choisi

Pour aboutir notre projet, nous avons dû employer un certain nombre d'outils et logiciels :

- **Microsoft Business Intelligence :**

Microsoft a montré un intérêt marqué pour la prise de décision ces dernières années, l'entreprise ciblant un segment significatif du marché, et avec les solutions qu'il propose, on peut dire qu'il est trajectoire correcte. La plate-forme Microsoft est livrée avec une variété d'outils qui vous permettent d'effectuer des performances élevées tout au long du processus décisionnel. Nous citons :

- SQL Server Intégration services (SSIS) ;
- SQL Server Analysis Services (SSAS) ;
- Microsoft Power BI

- **Système de gestion de base de données :**

SQL Server Management Studio (SSMS) est un outil pour accéder, configurer, Administration, administration et développement de tous les composants de SQL Server. Associés SSMS Un ensemble d'outils graphiques pour habilitier les développeurs avec un puissant éditeur de scripts et les administrateurs de tous les niveaux peuvent accéder à SQL Server.



FIGURE 2.10 – SQL Server logo

— **SSDT pour Visual Studio 2017 :**

SSDT est l'abréviation de Microsoft SQL Server Data Tools, c'est un outil de dévelop-



FIGURE 2.11 – Visual Studio logo

pement moderne et un complément à Visual Studio qui permet de développer un projet BI complet avec ces différentes phases.

Avec SSDt nous allons utiliser SQL Server Integration Services (SSIS) comme une solution d'intégration et de transformation des données et SQL Server Analysis Services (SSAS) comme une solution d'analyse des données.

— **Power BI Desktop :**

Power BI, Business Intelligence développée par Microsoft, est une suite d'outils d'informatique décisionnelle intégrée à Microsoft Office 365 pour :

- Traiter, analyser des données, provenant de multiples sources,
- Les transformer en informations exploitables, c.-à-d. les visualiser sous forme de rapports et tableaux de bord interactifs, en fonction de vos besoins descriptifs ou prédictifs,
- Les partager avec les équipes concernées, grâce à Office 365,
- Prendre des décisions tactiques ou stratégiques. [6]

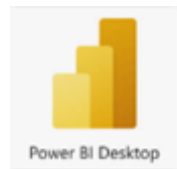


FIGURE 2.12 – Power BI Desktop logo

## 2.6 Démarches et Méthodologies

L'informatique décisionnelle est tout un concept, qui touche non seulement au côté technologique mais il met en valeur aussi l'importance du choix d'une méthodologie de travail. Le choix d'une méthodologie adaptée à la société est indispensable pour la réussite du projet. Il faut notamment prendre en compte la vitesse d'évolution des besoins et des priorités.

Bien que chaque projet de Business Intelligence (informatique décisionnelle) ou BI soit unique et réponde à des particularités techniques et processus d'exécution distincts, il est possible de définir quelques étapes ou phases d'un projet de BI, ainsi qu'une série de caractéristiques communes dans presque tous les cas.

- **Étape 1. Définition des objectifs et besoins**

Il s'agit là d'une phase essentielle du processus et elle se caractérise par la précision, car, outre la définition des besoins et objectifs, il faut déterminer quelles décisions concrètes doivent être prises, quel type d'informations est nécessaire et sur quelles variables l'analyse se basera-t-elle [2].

- **Étape 2. Choix de la méthodologie**

Toujours en fonction des objectifs et besoins du projet, il faut ensuite choisir la méthodologie concrète et les outils de BI à utiliser, ce qui implique aussi la formation et mise en place des équipes de travail [2].

- **Étape 3. Mise en place du programme de travail**

Il convient de définir, de manière détaillée, précise et claire, toutes les actions à réaliser, ainsi que l'infrastructure et les ressources nécessaires pour englober la méthodologie d'analyse de données choisie ainsi que les délais d'exécution du projet de Business Intelligence [2].

- **Étape 4. Actions de présentation**

L'étape suivante consiste à élaborer et présenter des rapports, reports, tableaux de bord, diagrammes de flux et infographies, de façon très visuelle, claire et schématique, afin de faciliter le travail des professionnels chargés de prendre des décisions??.

- **Étape 5. Exécution du système, formation et de support**

L'exécution d'un processus de Business Intelligence ne sera utile que si les informations, correctement analysées, arrivent aux personnes ayant la capacité de décider dans le support et avec les outils adéquats [2].

### 2.6.1 Méthodologie de travail

Pour réaliser et réussir un projet, il est important de suivre une méthodologie adaptée. Il s'agit d'un outil qui vous aide à accomplir votre projet étape par étape, de la planification à la mise en œuvre, dans un souci d'efficacité et rentabilité [3].

Il existe de nombreuses approches de la gestion de projet et faire le bon choix n'est pas toujours facile car il n'y a pas d'approche unique pour un projet, une entreprise ou une industrie. Une méthode fonctionne pour une situation mais pas pour une autre, et parfois vous devez combiner différents éléments de plusieurs méthodes pour créer la vôtre

. Choisir la bonne méthodologie de gestion de projet est donc une étape indispensable pour réussir votre projet. Optez pour une méthode qui valorise les objectifs de votre projet et les points forts de votre équipe. Vous pouvez également créer une méthodologie hybride qui assemble des éléments de différentes méthodologies. Traditionnelle, agile ou autres, l'essentiel est que vous et votre équipe trouviez la méthode de travail qui vous convienne le mieux.

### 2.6.2 Approches de modélisation BI Adopté

Pour faire une conception d'entrepôt de données (DWH), deux des approches d'entrepôt de données les plus largement discutées et expliquées sont les méthodologies Inmon et Kimball. Pendant des années, les gens ont débattu de l'approche d'entrepôt de données la meilleure et la plus efficace pour les entreprises. Cependant, il n'y a toujours pas de réponse définitive car les deux méthodes ont leurs avantages et leurs inconvénients.

#### 2.6.2.1 Inmon : de l'entrepôt aux magasins

Dans les années 1990, William Inmon propose une vision descendante de la structuration de la donnée de l'entreprise, dite "top-down".

L'entrepôt de donnée a pour rôle de rassembler toutes les données de l'entreprise, et les magasins de données sont alors des sous-parties issues de cet entrepôt.

#### 2.6.2.2 Kimball : des magasins à l'entrepôt

Ralph Kimball propose une vision radicalement différente, ascendante, de la structuration de la donnée, dite "bottom-up".

Les besoins de l'entreprise sont modélisés dans les magasins de données, l'entrepôt de données n'est par la suite qu'une simple union de ses magasins de données.

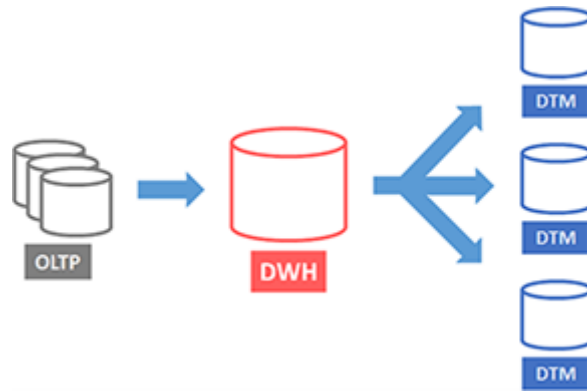


FIGURE 2.13 – L’approche top-down

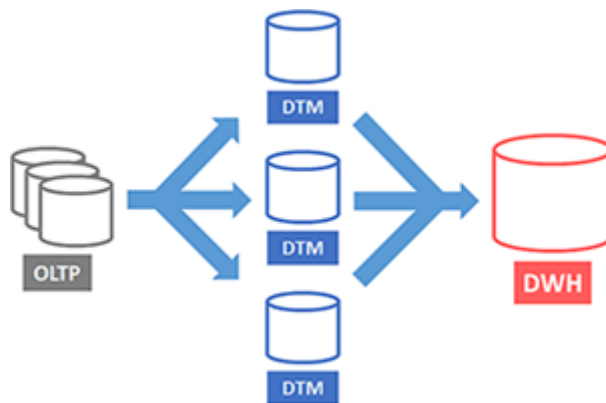


FIGURE 2.14 – Approche bottomup

D’après le cycle de vie de notre projet on a utilisé l’approche de Ralph KIMBALL qui nous permet de mettre en place une conception appuyée sur des phases de fonctionnement clairement établies et des points de vérification prédéfinis, avec un avantage du schéma en étoile est que la plupart des opérateurs de données peuvent facilement le comprendre en raison de sa structure dénormalisée, ce qui simplifie les requêtes et l’analyse.

## 2.7 Conclusion

Dans ce chapitre, j’ai commencé par une présentation des concepts de base des systèmes décisionnels (BI) et leurs divergences avec les systèmes transactionnels. Par la suite, on a détaillé l’architecture globale d’un cycle BI et en particulier ses différents composants. Puis on s’est intéressé aux différentes solutions BI présentes dans le marché. Pour en conclure, j’ai présenté les deux méthodologies celle de Inmon et l’autre de Kimball (utilisé vu les moyens, la cible et la durée de notre stage).

# Chapitre 3

## Conception et Développement de l'ETL

### 3.1 Introduction

L'étude de l'existant nous a permis de déceler les objectifs et les attentes de l'ISET et elle nous a fourni une vision claire de ce que nous allons aborder dans ce troisième chapitre. Dans ce chapitre, on va exposer l'architecture sur laquelle sera fondé notre système ce qui nous permettra de réaliser respectivement la modélisation, la conception et finalement l'implémentation du Data Warehouse.

Les étapes de l'entrepôt de données doivent être conçues de manière que les bonnes informations soient facilement accessibles au bon moment. L'architecture choisie est représentée par un modèle logique sous forme de 4 couches.

- Couche source de données ;
- Couche de préparation de données (Data StagingDataSource) ;
- Couche de stockage de données (DWH) ;
- Couche de présentation et restitution de données ;

On va détailler ces 4 couches selon notre sujet comme suit :

### 3.2 Couche source de données

Cette couche représente les différentes sources de données dans différents formats du système d'exploitation qui sont prêtes à être introduites dans l'entrepôt de données. Dans notre cas (sous forme de fichiers Excel)

Pour réaliser cette étape, il a fallu extraire les relations existantes entre les tables des bases de données, pour cela j'ai utilisé l'outil : MySQL Workbench pour faire les liaisons entre les tables et fournir une base de données plus au moins claire qui nous a donné le diagramme suivant :

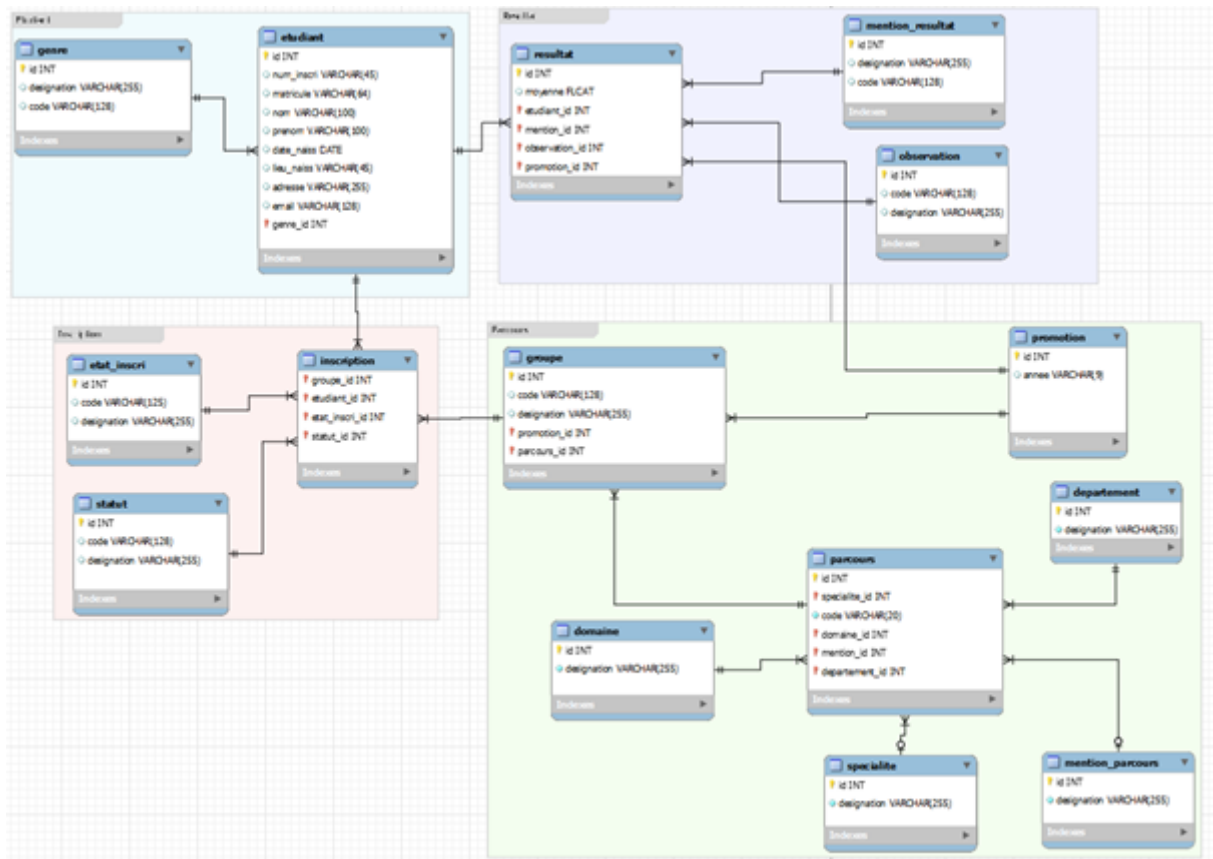


FIGURE 3.1 – Modèle Relationnel de la Base de données Partie relation entre les tables

### 3.3 Couche de préparation de données (StagingData-Source)

Cette couche est la couche de préparation des données (Data Staging), principalement pour une extraction rapide de Données provenant de sources d'information. Après avoir chargé les données Dans StagingDataSource, l'étape de transformation entre en jeu (suppression redondance des données, filtrage des données erronées) et Validez les données avant de les transférer vers DWH.

On peut nommer cette couche par la partie **ETL** (Extract-Tansform-LOAD).

#### 3.3.0.1 Extract : Extraction de données brutes

La première étape et la préparation de la base de données Staging avec la création de la table **audit\_logs** et les **fonctions**

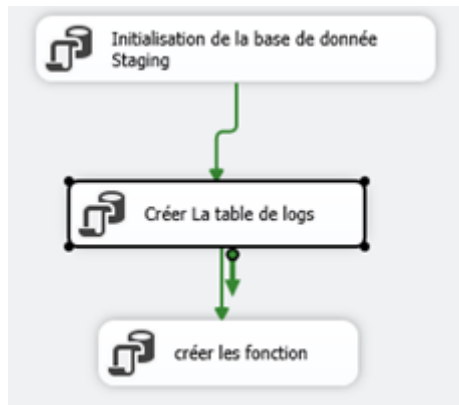


FIGURE 3.2 – Préparation au Staging

- **Audit<sub>logs</sub>** appelé aussi journal d'audit est essentiellement un enregistrement des événements et des modifications. Il nous permet de noter tous les changements au sein de notre travail, fournissant un historique complet des opérations réalisé.

SQLQuery4.sql - loc...G\91529\Fedia (75)) SQLQuery3.sql - loc...G\91529\Fedia (53)) SQLQuery2.sql - loc...G\91529\Fedia (65)) SQLQuery1.sql - loc...G\91529\Fedia (65))

```

/***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [Id]
      ,[PackageName]
      ,[TypeName]
      ,[TableName]
      ,[FileName]
      ,[RecordsInserted]
      ,[RecordsUpdated]
      ,[RecordsIgnored]
      ,[RecordsDate]
FROM [StagingDataSource].[dbo].[audit_log]
  
```

21 %

Results Messages

	Id	PackageName	TypeName	TableName	FileName	RecordsInserted	RecordsUpdated	RecordsIgnored	RecordsDate
1	1	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\20-05-2022-etudiants.csv	21391	0	289	2022-06-23 10:13:33.320
2	2	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\telechargement1.csv	1709	0	3	2022-06-23 10:13:33.823
3	3	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\telechargement2.csv	5413	0	77	2022-06-23 10:13:34.523
4	4	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 AA.csv	114	14	0	2022-06-23 10:13:35.037
5	5	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 FI.csv	32	1	0	2022-06-23 10:13:35.540
6	6	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 GC.csv	116	13	2	2022-06-23 10:13:36.050
7	7	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 GE.csv	276	34	18	2022-06-23 10:13:36.620
8	8	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 GM.csv	140	24	0	2022-06-23 10:13:37.143
9	9	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 INFO.csv	112	1	19	2022-06-23 10:13:37.673
10	10	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2015\L1 MKG.csv	89	12	0	2022-06-23 10:13:38.187
11	11	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2016\L1 AA.csv	185	16	0	2022-06-23 10:13:38.733
12	12	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2016\L1 FI.csv	60	0	0	2022-06-23 10:13:39.353
13	13	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2016\L1 GC.csv	154	5	0	2022-06-23 10:13:39.877
14	14	Etudiants	Import	ImportEtudiant	D:\pfe\import\etudiants\2016\L1 GE.csv	140	38	8	2022-06-23 10:13:40.543

FIGURE 3.3 – affichage de l'historique des opérations

- **Les fonctions** sont créées pour nous aider dans la phase de nettoyage et transformation des données.



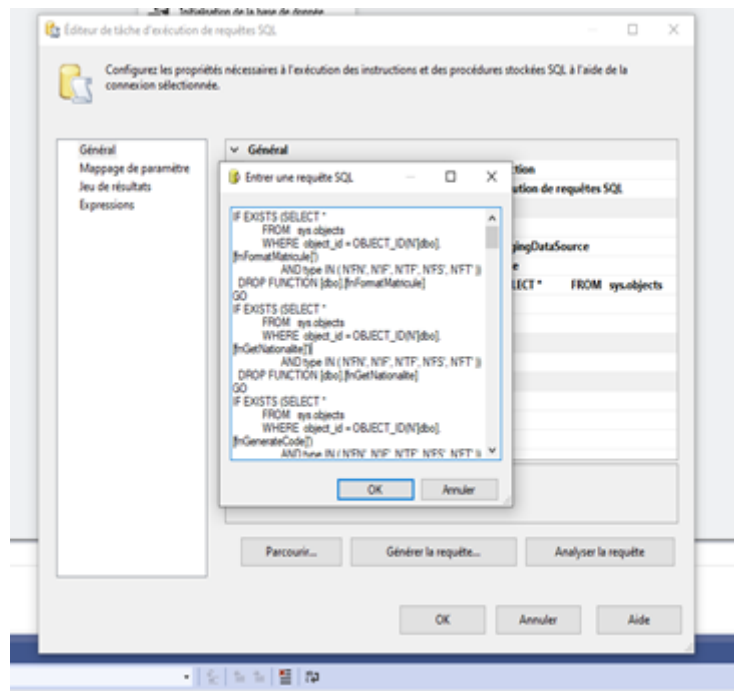


FIGURE 3.4 – création des fonctions

```
CREATE FUNCTION dbo.fnFormatMatricule ( @str VARCHAR(255))
RETURNS VARCHAR(50)
AS
BEGIN
    SET @str=RTRIM(LTRIM(@str));

    IF @str LIKE '[0-9]%'
    SET @str=CAST(FORMAT(CAST(@str AS NUMERIC), '0#####') AS VARCHAR(50));
    ELSE
    SET @str=CAST(LTRIM(RTRIM(@str)) AS VARCHAR(50));

    RETURN @str;
END
```

FIGURE 3.5 – création dbo.fnFormatMatricule

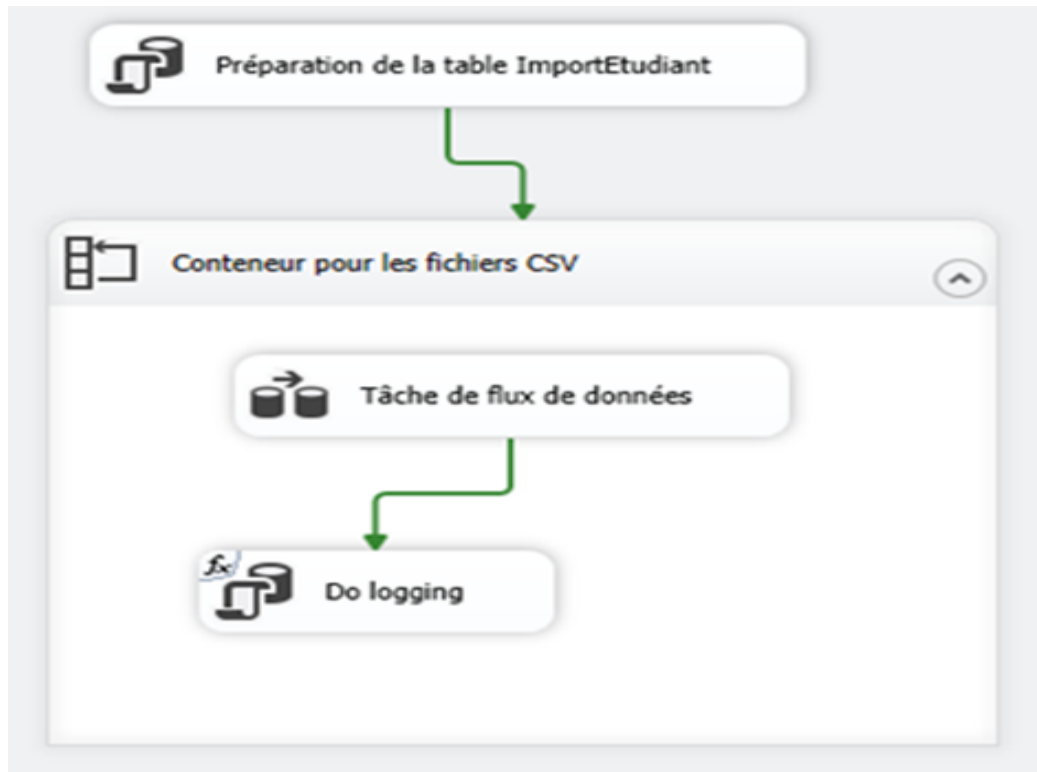


FIGURE 3.6 – Chargement de la table ImportEtudiant

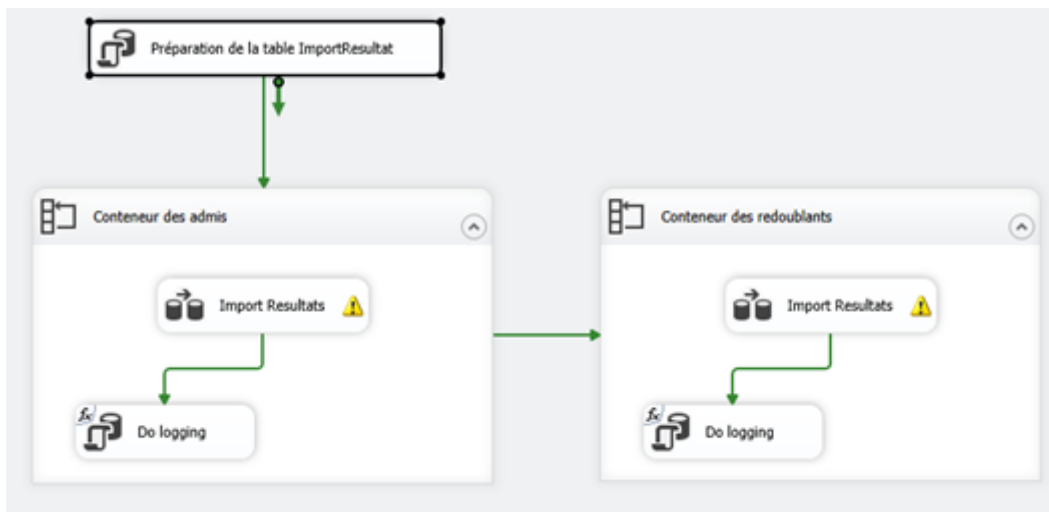


FIGURE 3.7 – Chargement de la table ImportResultat

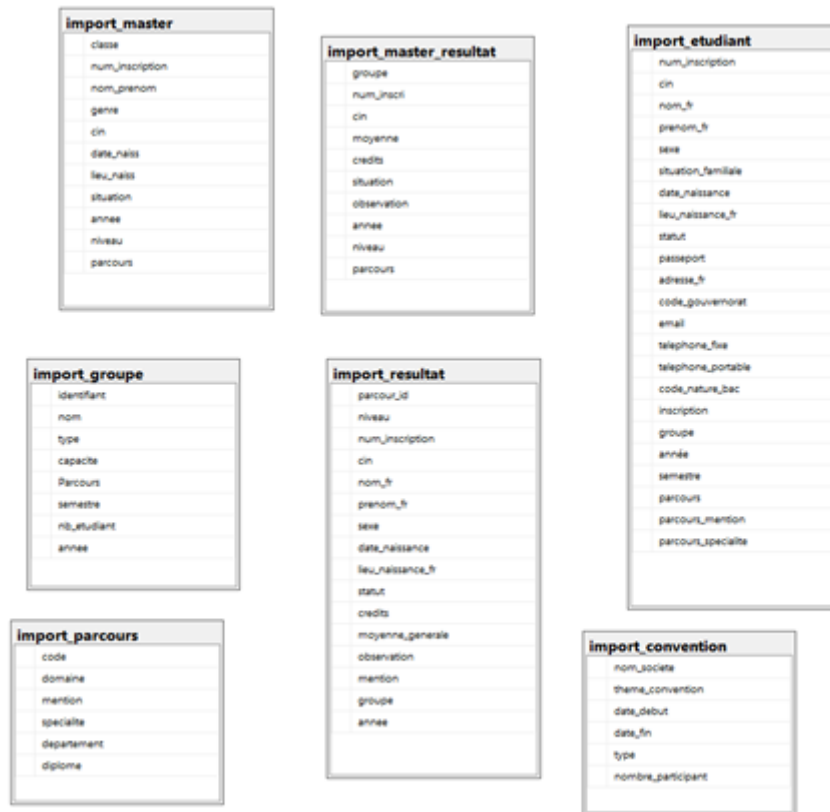


FIGURE 3.8 – Modèle Relationnel de la Base de données Partie ImportFichier

Le modèle ci-dessus est le résultat du processus d'import des fichiers Excel qui sont transformé en format CSV importé de SALIMA et du logiciel conçu pour les étudiants de maîtrise et un autre fichier Excel pour les conventions. Ce modèle se compose de 7 tables différentes, qui sont :

- **Table ImportMaster** : Contient les différentes informations sur un étudiants maitère (cin, numéro inscription, genre, nom prénom, année, niveau, parcours,etc) ;
- **Table ImportMasterResultat** : Contient les différentes informations concernant résultat Mastère (groupe, cin, numéro inscription, moyenne, crédit, observation, année, niveau, parcours, situation) ;
- **Table ImportEtudiant** : Contient les différentes informations concernant tous les étudiants ( cin, numéro inscription, nom, prénom, email, statut, état, année, semestre, parcours,...etc) ;
- **Table ImportGroupe** : Contient les différents groupes existant à l'ISSET( nom groupe, type, capacité, parcours, semestre, année) ;
- **Table ImportResultat** : Contient les différentes informations concernant résultat des étudiants( cin, numéro inscription, nom, prénom, moyenne général, observation,

mention, année,...etc);

- **Table ImportConvention** : Contient les différentes informations concernant les conventions( nom société, thème convention, type convention, date début, date fin, nombre participant);
- **Table ImportParcours** : Contient les différentes informations concernant les parcours existants( domaine, spécialité, département, diplôme);

### 3.3.0.2 Tansform : Transformation, homogénéisation et nettoyage de ces données



FIGURE 3.9 – Transformation et nettoyages de données pour la création de la table Etudiants

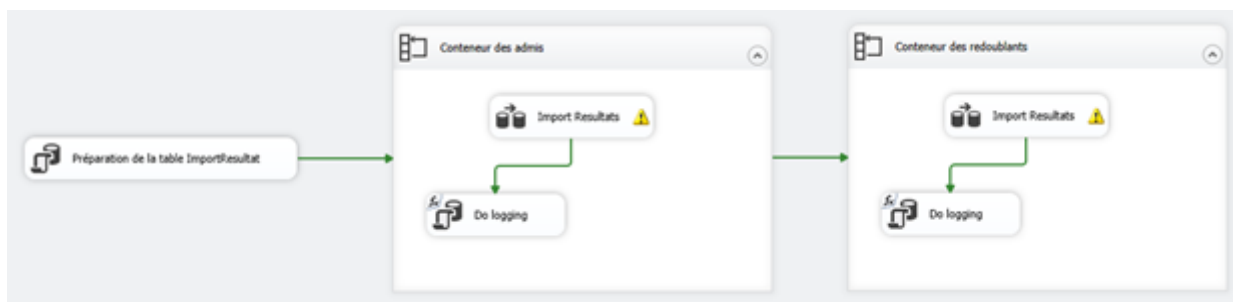


FIGURE 3.10 – Transformation et nettoyages de données pour la création de la table Resultat

Ces 2 modèles ci-dessus est le résultat de la transformation des données à partir des tables (Table **ImportMaster**, Table **ImportMasterResultat**, Table **ImportEtudiant**, Table **ImportGroupe** , Table **ImportResultat**, Table **ImportConvention**, Table **ImportParcours**). Ce modèle se compose de 20 tables différentes, qui sont :

- Table Etudiant.



FIGURE 3.11 – Transformation et nettoyages de données pour la création de la table Inscription



FIGURE 3.12 – Modèle Relationnel de la Base de données Partie Etudiant

- Table Inscription.
- Table Parcours.
- Table Nature Bac.
- Table Genre.

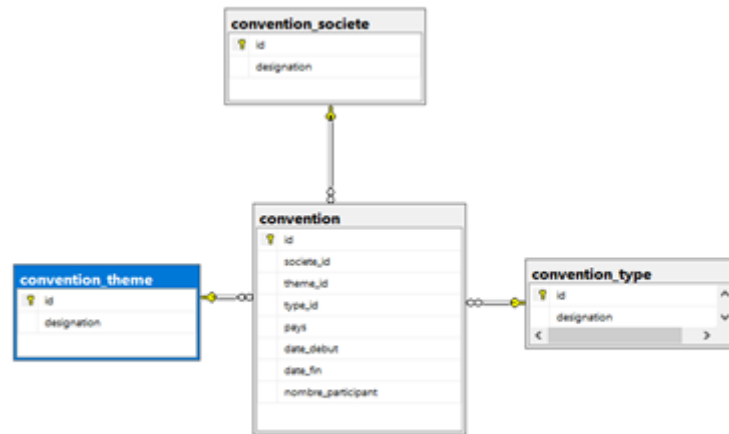


FIGURE 3.13 – Modèle Relationnel de la Base de données Partie Convention

- Table Gouvernorat.
- Table Inscription Etat.
- Table Inscription Statut.
- Table Groupe.
- Table Résultat.
- Table Département.
- Table Résultat Mention.
- Table parcours Diplôme.
- Table parcours Spécialité.
- Table parcours Département.
- Table parcours Domaine.
- Table convention.
- Table Convention Type.
- Table Convention Thème.
- Table Convention Partenaire.

### 3.4 Couche de stockage de données (DWH)

C'est la couche de l'alimentation de l'entrepôt de données (DWH) pour stocker des informations analytiques et historiques qui seront utilisées.

Après la transformation et le nettoyage des données et les stocker dans notre Staging on passe maintenant à l'alimentation des tables de fait et dimension et on a obtenu ce diagramme :



FIGURE 3.14 – Modèle en étoile (DWH)

### 3.5 Couche de présentation et restitution de données

Cette couche fait référence aux informations qui parviennent à l'utilisateur final, telles que les rapports, grâce à cette couche, l'utilisateur pourra :

- Analyser les données, notamment avec les outils de type OLAP pour les analyses multidimensionnelles
- Aide à la décision des décideurs en situation avec les tableaux de bord présentant les indicateurs clés de l'activité
- Transmettre l'efficacité avec le Reporting.

Cette partie sera plus détaillée dans le chapitre 4

# Bibliographie

- [1] BI. Developpement-dune-solution-business. <http://dspace.univ-tlemcen.dz/bitstream/112/10989/1/Developpement-dune-solution-Business.pdf>, 2022.
- [2] CAPTIO. projet-de-business-intelligence. <https://www.captio.fr/blog/en-quoi-consiste-un-projet-de-business-intelligence>, 2022.
- [3] CHOISIR METHODOLOGIE. etapes-choisir-methodologie. <https://www.planzone.fr/blog/etapes-choisir-methodologie-gestion-projet>, 2022.
- [4] IRIT. Rapport ben kraiem ines. [https://www.irit.fr/neocampus/images/stages/2017/2017\\_Rapport\\_BEN\\_KRAIEM\\_Ines.pdf](https://www.irit.fr/neocampus/images/stages/2017/2017_Rapport_BEN_KRAIEM_Ines.pdf), 2022.
- [5] KEMP, S. a short history of datawarehousing. <http://www.dataversity.net/a-short-history-of-data-warehousing/>, 2022.
- [6] POWER BI. nos-metiers. <https://www.groupe-hli.com/nos-metiers/power-bi/>, 2022.
- [7] SONAR. sonar. <https://sonar.test.rero.ch/documents/54/preview/BusinessIntelligence.pdf>, 2011.
- [8] TECHDIFFERENCES. fact-and-dimension-table. <http://techdifferences.com/difference-between-fact-and-dimension-table.html>, 2022.
- [9] WIKIPEDIA. Datawarehouse. [https://en.wikipedia.org/wiki/Data\\_warehouse](https://en.wikipedia.org/wiki/Data_warehouse), 2022.



**Annexe A**

**Annexes**