



République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université des Sciences et de la Technologie Houari Boumediene



Faculté d'Electronique et d'Informatique
Département Informatique

Mémoire de Licence

Filière: Informatique

Spécialité: ISIL

Thème

Conception et mise en place d'un entrepôt de données pour gérer l'activité du département informatique par une méthode de modélisation Faits/ Dimensions

Soutenu le :.../.../....

Sujet Proposé et encadré par :

Mme HANK Karima

Présenté par :

Boulmaali Linda Imene

Aissani Anouar

Devant le jury composé de :

M..... Président (e)

M..... Membre

Binôme N° : ISIL 042/21

Remerciements

En préambule à ce mémoire de licence, nous tenons avant tout à remercier Dieu de nous avoir donné le courage et la force pour mener à bien ce projet.

Aussi, nous tenons à remercier notre encadrante Mme HANK Karima de nous avoir proposé ce sujet, qui à travers, nous avons pu apprendre pleins de choses.

Nous remercions toutes les personnes qui ont contribué de près ou de loin à la finalisation de notre projet.

On remercie vivement Messieurs les membres du jury d'avoir accepté d'évaluer ce travail.

Linda et Anouar

J'adresse mes plus sincères remerciements à mes parents, mon frère Ayoub et ma sœur Lilia, pour leur aide inestimable et leur soutien indéfectible.

Je remercie mon binôme Anouar AISSANI, pour sa patience et son soutien.

Et enfin, je remercie mes amis Assia, Meriem et les membres du club Open Minds pour leur soutien moral et encouragement qui m'a toujours poussé à donner le meilleur de moi même.

Linda

J'adresse mes plus sincères remerciements à mes parents, mes frères et sœurs spécialement ma grande sœur Wafa et ma petite soeur Baraa, ainsi que tous les membres de ma famille.

Je remercie mon binôme Linda Imene BOULMAALI, pour son sérieux et son soutien.

Et enfin, je remercie tous mes amis Yasmine, Karima, Lydia et Assia.

Anouar

Sommaire

Introduction générale

Contexte général	1
Problématique	2
Objectif du projet	2
Planification et conduite du projet	3

Chapitre I: Etat de l'art

Introduction	4
I.1. Les systèmes décisionnels:	4
I.2. Décisionnel vs transactionnel	5
I.3. Les domaines d'application du décisionnel	5
I.4. Le Data Warehouse	5
I.4.1. Qu'est ce qu'un Data Warehouse	5
I.4.2. Historique des Data Warehouse	6
I.5. Modélisation dimensionnelle d'un entrepôt de données (MD)	6
I.5.1. Concepts de la modélisation dimensionnelle	6
I.5.2. Différents modèles de la modélisation dimensionnelle	7
I.6. Eléments et Architecture d'une Data Warehouse	9
Conclusion	10

Chapitre II: Analyse de l'existant et des besoins

Introduction	11
II.1. Historique du département informatique	11
II.2. Organigramme du département	11
II.3. Analyse de l'existant	12
II.3.1. L'informatique au sein du département	12
II.3.2. Analyse des données du département	12
II.4. Analyse des besoins	13
II.4.1. Identification des acteurs	13
II.4.2. Liste des besoins	13
Conclusion	14

Chapitre III: Conception de la solution

Introduction	15
II.1. La base de données relationnelle	15
II.2. La conception du datawarehouse	15
II.2.1 Justification de notre approche	15
II.2.2. La modélisation dimensionnelle	16
II.2.2.1 Datamart « StudyEnseignement »	16
II.2.2.2. Datamart « StudyEtudiant »	17
II.3. Alimentation du Data Warehouse	18

II.3.1. L'extraction des données	18
II.3.2. Le chargement des données	18
II.4. Diagramme des cas d'utilisation	19
Conclusion	20
Chapitre VI: Réalisation et test	
Introduction	21
VI.1. Environnement de travail	21
VI.1.1 Outils technologiques utilisés	21
VI.1.2. Langages utilisés	22
VI.2. Présentation de l'application	22
Conclusion Générale	28
Bibliographie	29
Webographie	29
Liste des figures	30
Liste des tableaux	30
Annexe	
Résumé	

Introduction générale

Contexte général

Nous vivons dans une ère d'explosion quantitative de données numériques. Ces données ont un usage essentiel au sein d'une organisation, et peuvent être exploitables de manière à prendre les bonnes décisions pour pouvoir atteindre les objectifs et intégrer le marché de la concurrence.

En effet les dirigeants de n'importe quel établissement quelque en soit d'ailleurs le domaine d'activité, doivent être en mesure de mener à bien les missions qui leur incombent en la matière. Ils devront prendre notamment les décisions les plus opportunes, la supervision et l'anticipation des comportements à venir au sein de l'établissement. Ces décisions sont prises par l'évaluation des stratégies de l'organisation à travers des chiffres et des informations fiables et concrètes ainsi que l'analyse prédictive à travers des algorithmes statistiques.

Cependant, quel que soit l'utilité de ses données au sein d'une entreprise, celles-ci doit pouvoir être manipulable et exploitable facilement pour apporter de la valeur à l'établissement, car très souvent l'architecture de ce dernier rend les données potentiellement utiles inaccessibles, cela est dû à la masse considérable de données que peut contenir un établissement ainsi que la diversité des sources d'en elles proviennent parfois même hétérogènes.

C'est dans ce contexte que de nouvelles méthodes dont les « **entrepôts de données** » (Data Warehouse) ont vu le jour. Ces technologies représentent l'élément principal et incontournable pour la mise en place d'un bon système décisionnel.

De part sa dimension éducative et sa position au sein de l'université des sciences et de la technologie Houari Boumediene, les bases de données du département informatique contiennent un volume important de données en croissance continue, ces données proviennent de sources hétérogènes. L'élaboration d'un système décisionnel est donc nécessaire pour homogénéiser ces données et mettre à la disposition de la direction du département Informatique un outil d'aide à la décision en s'appuyant sur l'analyse des indicateurs pertinents.

Notre projet de fin d'étude tend à la mise en place d'un système en mesure d'exploiter les données issues des systèmes transactionnels, et d'offrir des informations de qualité pour les décideurs. Un tel système requiert la mise en place d'un entrepôt de données fiables contenant les informations nécessaires à l'accomplissement des processus décisionnels.

Problématique

Le département informatique de l'USTHB rattaché à la faculté d'électronique et informatique, recueille une quantités considérable de données recouvrant différents parcours dont 3 licences et 7 masters, sans oublier les laboratoires de recherche informatique qui rendent la manipulation des données et la tâche administrative plus ardues.

La prise de décision stratégique dans un tel établissement nécessite le recours et le croisement de multiples informations qui concernent tous les services du département, inscription, délivrance des diplômes, organisation de l'emploi du temps et des examens, le suivi individualisé des étudiants par filière, et aussi le suivi des déroulements des formations.

Or les données collectées par le département informatique ne sont pas exploitées de la meilleure manière à permettre une analyse fiable sans des moyens considérables sur des périodes plus ou moins longues. Ainsi, les principales difficultés rencontrées peuvent être résumées en :

- La difficulté de rassembler des données éparpillées non connectées entre elles au sein des services du département, et l'hétérogénéité de ces données dans leurs formats techniques et leurs organisations structurelles, ce qui amène à une procédure archaïque, lente avec d'éventuelles erreurs.
- Les données sont implémentées pour l'action (par construction) et non pour l'analyse, on constate de ce fait une insuffisance des états statistiques permettant l'aide à la prise de décision.
- La procédure du reporting au niveau de l'établissement est assez coûteuse et cela est principalement dû au nombre d'intervenants et des moyens mis en place pour cette dernière.
- Les données peuvent être volatiles, au sens où leur mise à jour peut conduire à oublier des informations obsolètes.

Contexte du projet

1. Objectif du projet

Ce projet a pour but de pallier les problèmes précédemment cités, en mettant en disposition du département informatique un entrepôt de données contenant les informations nécessaires ayant pour objectifs de:

- L'amélioration du pilotage interne de l'établissement, en vue d'aider à la prise de décision.
- L'amélioration du dialogue de gestion en matière d'orientation ou de pédagogie.

- Offrir des informations fiables et pertinentes, permettant aux décideurs et aux analystes la possibilité de faire des analyses appropriées.
- La réduction du nombre d'intervenants lors de la production de rapports.

2. Planification et conduite du projet

Afin de mener notre projet à bien et contenu de la date limite, nous avons suivi un plan, qui nous a permis de définir les tâches réalisées, et d'avoir un aperçu sur l'état d'avancement.

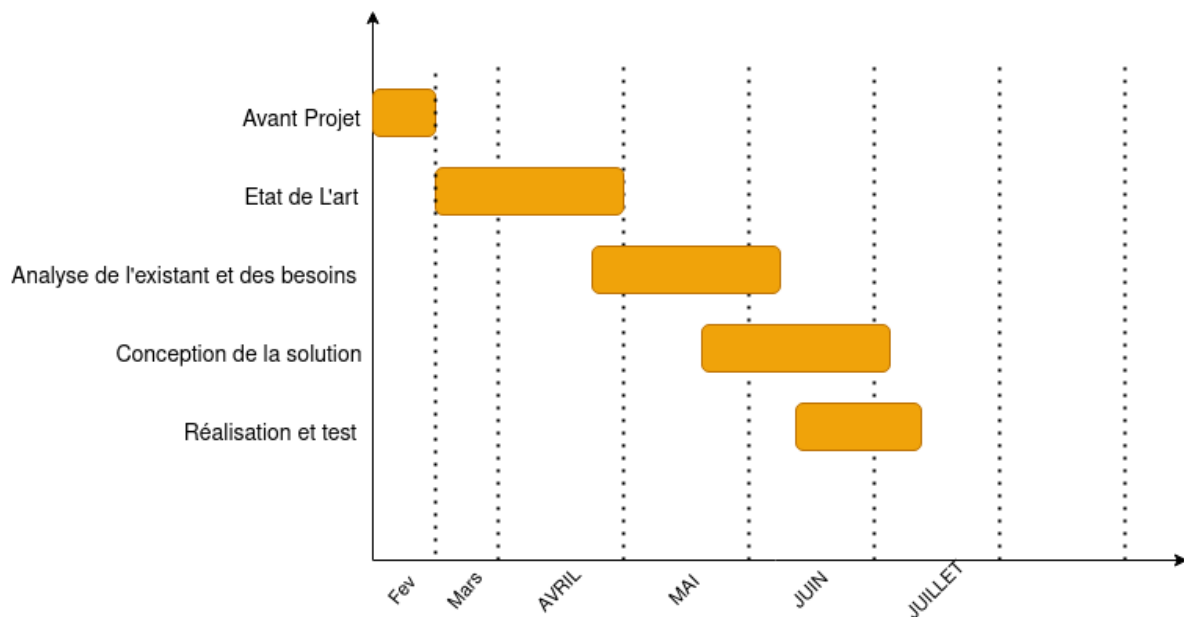


Figure 1: Diagramme de Gantt de l'avancement du projet

Le présent rapport de projet de fin d'étude s'articule autour de 4 chapitres :

Chapitre 1 : est destiné à la présentation des concepts de base de l'informatique décisionnelle ainsi que les entrepôts de données.

Chapitre 2 : est consacré à l'étude de l'existant et la définition des besoins.

Chapitre 3 : est consacré à la conception de la solution proposée.

Chapitre 4 : est le dernier chapitre qui décrit toutes les phases du processus de réalisation de l'application.

Nous terminerons ce document par une conclusion générale qui résume le travail effectué avec des perspectives pour des futurs travaux en relation avec ce projet.

Chapitre I: Etat de l'art

Introduction:

L'entrepôt de données (Data Warehouse), constitue une structure informatique pour la mise en place d'applications décisionnelles.

Dans ce chapitre nous visons à donner un aperçu général sur les systèmes d'aide à la prise de décision et les outils qui les composent.

I.1. Les systèmes décisionnels:

Avant d'entamer la partie conception d'un Data Warehouse, nous allons définir quelques concepts clés autour du décisionnel puisque la construction d'un Data Warehouse conduit à la mise en place d'un système décisionnel dans une entreprise. Parmi les différentes définitions du BI (*Business Intelligence*) qui ont été données on trouve :

"Le Décisionnel est le processus visant à transformer les données en informations et, par l'intermédiaire d'interrogations successives, transformer ces informations en connaissances." [Dresner, 2001].

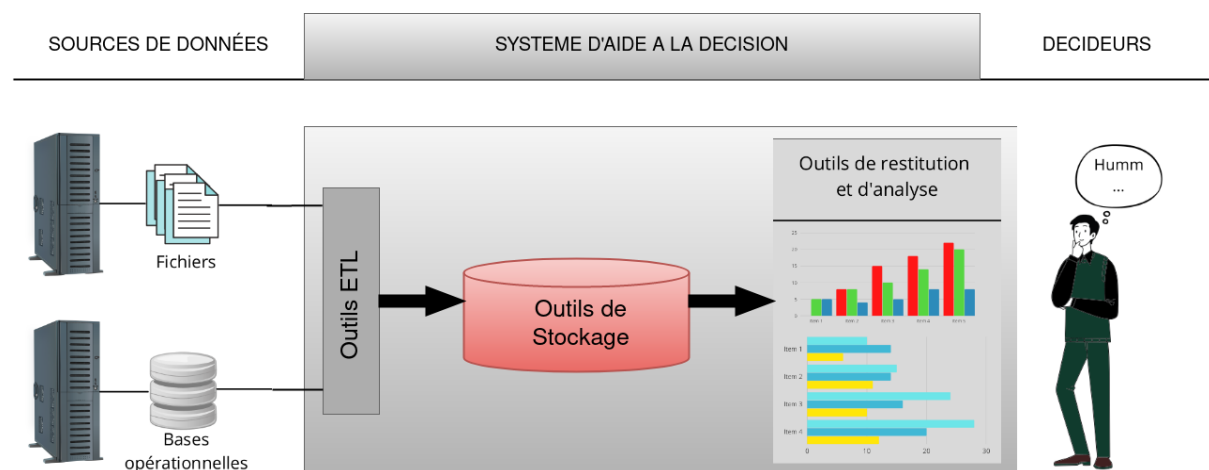


Figure I.1 -Le système d'aide à la décision.

Cette figure illustre trois catégories d'outils employés:

- Les outils d'extraction, de transformation et de chargement (ETL) des données opérationnelles pour alimenter et rafraîchir les données contenues dans le système d'aide à la décision.
- Les outils de stockage et de traitement des données décisionnelles, extraites des données opérationnelles.
- Les outils de restitution et d'analyse des données décisionnelles sous une forme adaptée aux décideurs.[1]

I.2. Décisionnel vs transactionnel

Le système de traitement transactionnel ou STT est une base de données classique destinée à assumer des **transactions** en temps réel : Ajout, mise à jour, suppression de données, questions sur des données identifiées ou questions statistiques., tandis que le système décisionnel est conçu dans le but d'extraire des données nécessaire à partir des base de données classique qui sert à faire des calculs et étudier des statistiques pour prendre des décisions.

I.3. Les domaines d'application du décisionnel

- La gestion de la relation client (CRM) est l'un des premiers champs d'application de la Business Intelligence.
- La direction commerciale pour le pilotage des réseaux, les prévisions des ventes, l'optimisation des territoires...
- Les ressources humaines pour la gestion des carrières,
- La direction de la production pour l'analyse qualité, la prévision des stocks, la gestion des flux, la fiabilité industrielle...
- La direction générale pour les tableaux de bord, indicateurs de pilotage, gestion d'alertes...

I.4. Le Data Warehouse

I.4.1. Qu'est ce qu'un Data Warehouse

Dans son livre "building the data warehouse"[Inmon, 2002], maintenant référence du domaine, Bill Inmon définit le Data Warehouse de la manière suivante:

« Le Data Warehouse est une collection de données orientées sujet, intégrées, non volatiles et évolutives dans le temps, organisées pour le support d'un processus d'aide à la décision. »

Les paragraphes suivants illustrent les caractéristiques citées dans la définition d'Inmon.

Orienté sujet : Le DW (*Data Warehouse*) est organisé autour des sujets majeurs de l'entreprise ou les données sont destinées à un processus analytique.

Intégrées : les données proviennent de différentes sources, de la plupart ou la totalité des applications de l'organisation.

Évolutif dans le temps : toute modification apportées aux données dans la base est suivie et enregistrée, cela permet le suivi de l'évolution des valeurs dans le temps.

Non volatile : une donnée dans un environnement DW ne peut être mise à jour ou supprimée contrairement à celle d'un environnement opérationnel.

I.4.2. Historique des Data Warehouse

Les systèmes décisionnels constituent l'élément déclencheur du concept « Data Warehouse ». En effet, cette technologie remonte aux années 80, durant lesquelles un intérêt croissant au système décisionnel a vu le jour, dû essentiellement à l'émergence des SGBD relationnels avec leur simplicité et la puissance offerte par le langage SQL.

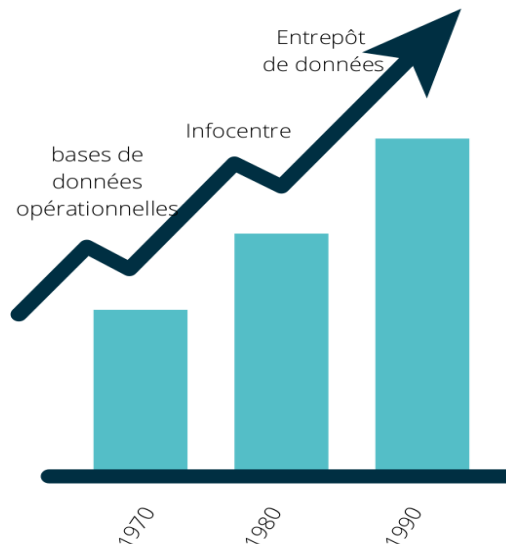


Figure I.2 -évolution des bases de données décisionnelles.

I.5. Modélisation dimensionnelle d'un entrepôt de données (MD)

La MD est une structure de données optimisée permettant le stockage de données dans un data warehouse d'une manière plus rapide et simple à comprendre pour offrir une bonne performance des requêtes aux utilisateurs, elle s'appuie sur les deux concepts suivants:

I.5.1. Concepts de la modélisation dimensionnelle

a). Concept de faits:

Une table des faits est la table primaire d'un modèle dimensionnel. Cette table stocke sur plusieurs lignes, des clés étrangères qui sont des clés primaires des tables de dimension, et des mesures qui donnent généralement les valeurs numériques du fait. Elle est considérée comme une table de jointure en rationnel qui assure des relations de type plusieurs à plusieurs entre les tables de dimensions.

«Les tables de faits expriment les relations multiples entre les dimensions dans les modèles dimensionnels. »[Kimball, Ross, 2008, p.19]

b). Concept de dimensions:

Ce sont des tables qui offrent les caractéristiques descriptives des faits à l'aide des attributs hautement corrélés, elle permet de faire des analyses des données pour que l'entrepôt de données devienne utilisable et compréhensible tel que chaque dimension peut contenir une ou plusieurs relations hiérarchiques.

« Une table de dimension établit l'interface homme / entrepôt, elle comporte une clé primaire » [Kimball, 2002].

c). Comparaison entre les tables de faits et les tables de dimensions:

Table de faits	Table de dimensions
Peu d'attributs	Beaucoup d'attributs
Beaucoup de données	Peu de données
Valeurs des mesures	Valeurs de description
Des clés étrangères	Une clé primaire
Augmente verticalement	Augmente horizontalement
Utiliser pour faire des calculs	Utiliser pour créer des contraintes
Moins de tables un schéma	Plus de tables dans un schéma

Tableau I.1 : Tableau comparatif entre les tables de faits et les tables de dimensions.

I.5.2. Différents modèles de la modélisation dimensionnelle

Tout comme une base de données, un entrepôt de données nécessite également de maintenir un schéma. Une base de données utilise un modèle relationnel, tandis qu'un entrepôt de données utilise le schéma en **étoile**, **flocon** ou **constellation**.

a). Le modèle en étoile:

Le schéma en étoile est considéré comme le bloc de construction de base dans la modélisation dimensionnelle, il se compose d'une table centrale appelée la **table des faits**, et un certain nombre de tables plus petites appelées **tables de dimension** qui rayonnent hors de la table centrale (une seule table pour chaque dimension).

“Le modèle en étoile est une représentation fortement dénormalisée qui assure un haut niveau de performance des requêtes même sur de gros volumes de données.” [2]

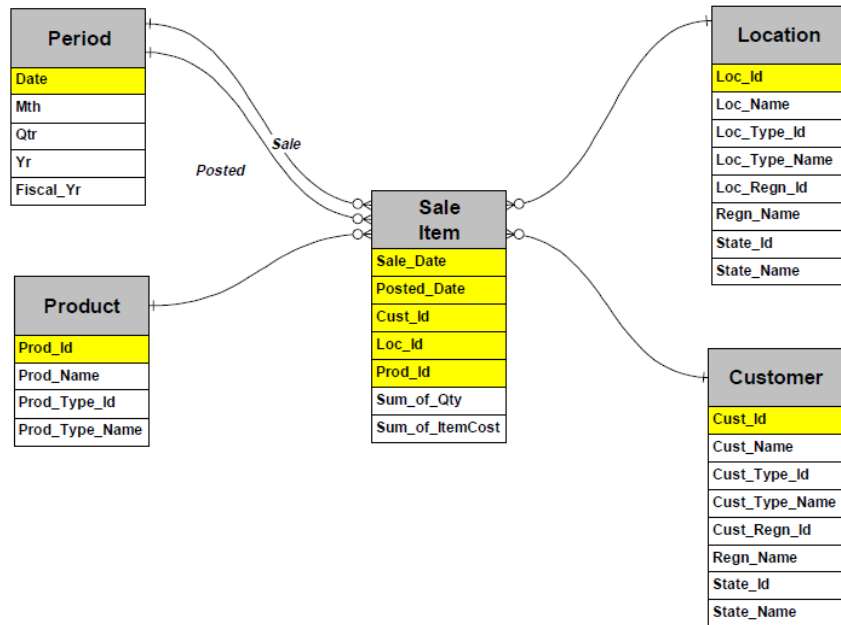


Figure I.3 : Schéma en étoile de “Sale Item”. [3]

b). Le modèle en flocon:

“Le modèle en flocon est aussi un modèle dénormalisé, mais un peu moins que le modèle en étoile : il conserve un certain niveau de décomposition pour chaque dimension prise isolément.” [2]

Le schéma en flocon est une variation du schéma en étoile qui a plusieurs niveaux de tables de dimensions normalisées.

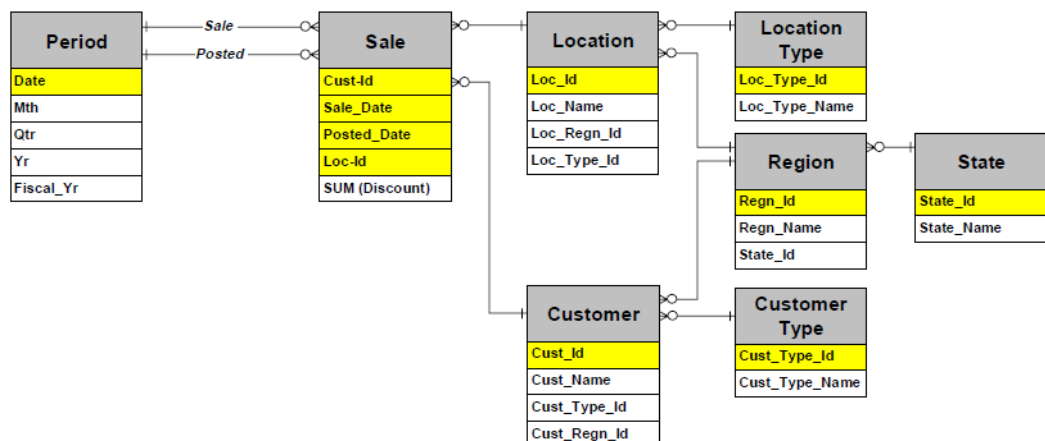


Figure I.4 : Schéma en flocon de “Sale”. [3]

c). Le modèle en constellation

Au lieu d'un certain nombre de schémas d'étoiles discrètes, le modèle de données peut être transformé en un schéma de constellation. Un schéma de constellation se compose d'un ensemble des schémas d'étoiles avec des tables de faits hiérarchiquement liés.

Un schéma de constellation se compose d'un ensemble des schémas d'étoiles avec des tables de faits hiérarchiquement liés.[2]

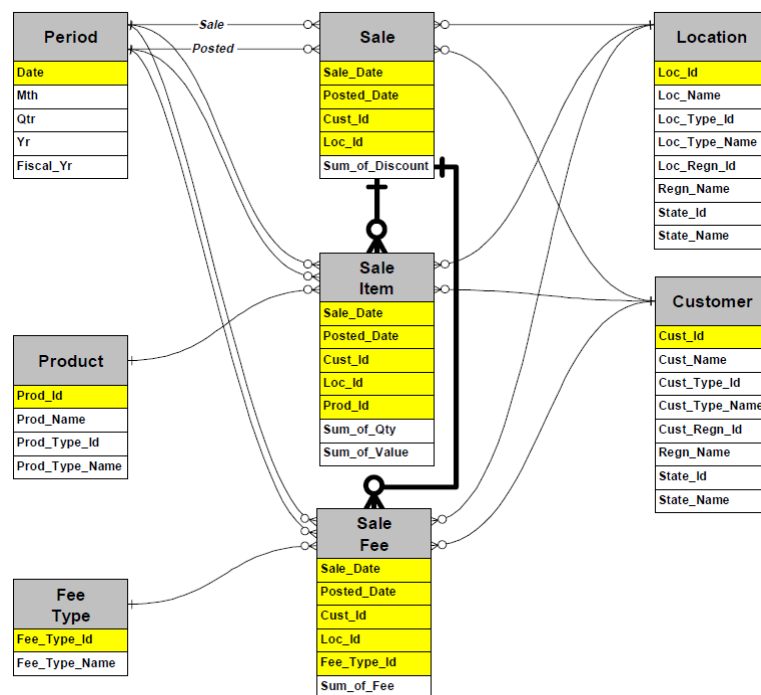


Figure I.5 : Schéma en constellation de "Sale".[3]

I.6. Eléments et Architecture d'une Data Warehouse

L'environnement du Data Warehouse est constitué essentiellement de quatre composantes :

Les applications opérationnelles: Ce sont des applications du système opérationnel de l'entreprise extérieures au Data Warehouse.

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

- Extraction des données primaires (issues de différentes sources),
- Transformation des données,
- Le chargement des données traitées dans l'entrepôt de données,

Ces trois étapes décrivent une mécanique cyclique qui a pour but de garantir l'alimentation du Data Warehouse en données homogènes, propres et fiables.[4]

Présentation des données: C'est l'entrepôt où les données sont organisées et stockées.

On distingue deux architectures internes du Data Warehouse :

1. Data Mart indépendant

Les Datamart sont des versions miniaturisées du Data Warehouse au niveau départemental, alimentées par le Data Warehouse et basées sur les besoins départementaux en informations [Inmon, 2002]

2. Data Mart interconnectés

Les Datamart sont construits autour de sujets, interconnectés grâce aux tables des faits contenues dans le Data Warehouse, ce dernier se compose alors des Datamart et ces tables des faits, appelées bus[Kimball, 2002].

Zone d'outils d'accès: Outil analytique pour la visualisation des données, tels que des tableaux de bord, indicateurs clés de performance ou Key performance Indicators (KPI), des rapports consolidés.

Conclusion

Afin de mettre en place une telle application, apte à l'aide à la prise de décision en sein du département informatique, il est nécessaire de choisir et d'adopter une démarche précise qui doit tenir compte des réalités de l'établissement et des contraintes du projet.

La modélisation de l'entrepôt se fait dans tous les cas grâce à la modélisation dimensionnelle. Au cours du 2eme chapitre de notre projet, nous allons essayer d'utiliser les concepts présentés dans la synthèse bibliographique, et cela afin de mettre en œuvre notre datawarehouse.

Chapitre II:
Analyse de l'existant et
des besoins

Introduction:

Dans ce chapitre nous visons à donner un aperçu sur l'organisme d'accueil et analyser les ressources existantes dans ce dernier, afin de pouvoir comprendre et de prévoir les éventuelles fonctionnalités dont les décideurs ont besoin pour une bonne prise de décision.

II.1. Historique du département informatique

Avant de passer à l'étude et l'analyse du département, nous allons faire un arrêt sur son historique.

Le département informatique a été créé en 1974 au sein de l'institut de mathématiques, sous l'appellation de DILM (département d'informatique et logique mathématique). Il forme alors les premiers ingénieurs d'état en informatique de l'USTHB.

En 1986, il devient institut d'informatique, pour se consacrer davantage à sa vocation première, la recherche dans le domaine de l'informatique et la formation d'ingénieurs d'état dans la même discipline.

Depuis l'an 2000, le département informatique revient avec pour objectif de former des informaticiens maîtrisant techniques et concepts fondamentaux, leur permettant de poursuivre un enseignement post-gradué et ou de se verser dans les domaines clés du marché de l'informatique.[5]

II.2. Organigramme du département

Comme nous l'avons mentionner dans la problématique de notre projet, le département informatique occupe plusieurs services et posts qu'on peut résumer dans l'organigramme suivant:

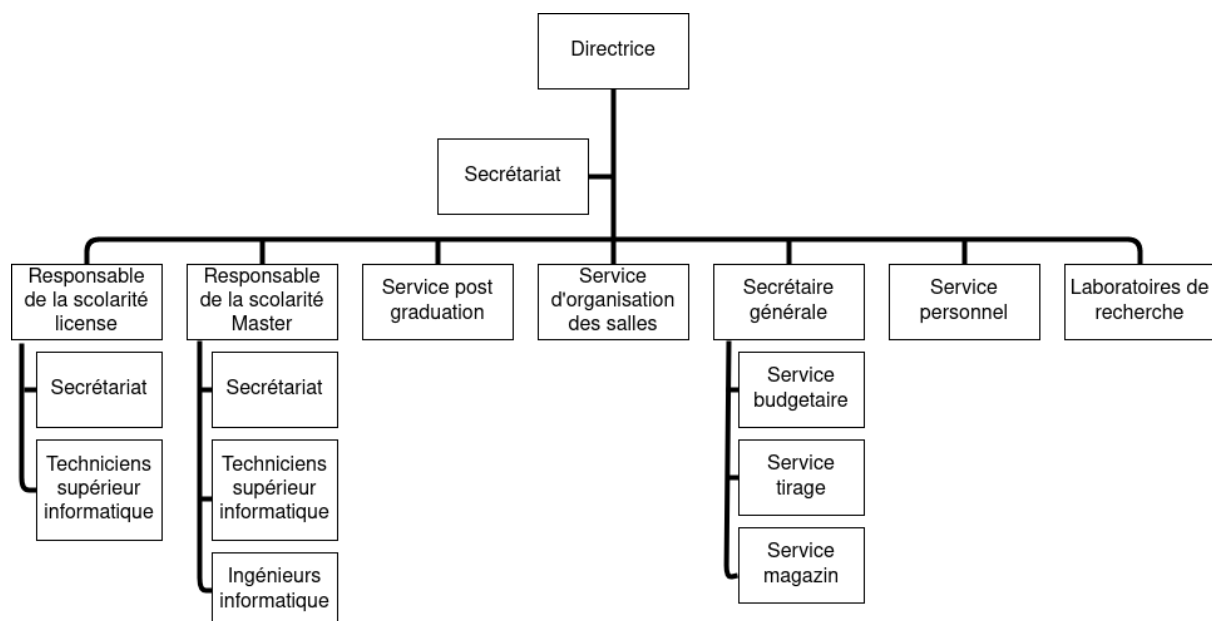


Figure II.1 : Organigramme du département informatique[6]

II.3. Analyse de l'existant

Après avoir donné une rétrospective du département informatique, nous allons maintenant nous intéresser à l'étude de l'environnement et analyser les structures des données des logiciels de gestion de scolarité.

II.3.1. L'informatique au sein du département

Des questions ont été posées sur les logiciels informatiques utilisés au niveau du département, La synthèse des résultats permet d'établir les constats suivants :

Les services de scolarité ainsi que la majorité des structures se suffisent à utiliser des logiciels de bureautique comme le MS-Office ou encore OpenOffice pour certains, ainsi que le logiciel **S.E.E.S** fourni par le ministère de l'enseignement supérieur pour la gestion des inscriptions, suivi des notes, etc., des étudiants durant leur cursus académique. Ce logiciel fonctionne sur une base de données Microsoft Access.[7]

Nous avons constaté que mis à part les logiciels de gestion, les services du département ne disposent pas de logiciels qui aident à la prise de décision.

II.3.2. Analyse des données du département

La collecte des données s'est effectuée à travers des sources hétérogène dont, les mémoires, et les tableaux excel trouvés sur internet. Nous nous sommes concentrés lors de notre recherches sur des données opérationnelles propres aux scolarités en générale afin que nous puissions satisfaire l'objet de notre étude. Voici un aperçu des données collectées.

Nom de données	Code de donnée	Type de donnée	Taille de donnée
matricule	matEtu	numérique	12
nom	nomEtu	chaîne	50
prenom	prenomEtu	chaîne	50
date de naissance	dateNai	date	10
lieu de naissance	lieuNai	chaîne	50
téléphone	ntph	numérique	10
Email	email	alphanumérique	250
Adresse	adresse	alphanumérique	250
Ville	ville	chaîne	50
Sexe	sex	chaîne	1

Tableau II.1 : Tableau des données de la table etudiant

Vu le nombre des tables trouvés, on a présenté qu'un seul exemple, le reste des tables sont reportées dans l'Annexe2.

II.4. Analyse des besoins:

Dans l'étude de l'existant nous avons pu avoir un aperçu sur les éventuels besoins du département, mais cela n'est guère suffisant pour avoir une idée claire sur la manière de construire notre datawarehouse ou de son architecture, la conception et la réalisation de ce dernier doit être en mesure de répondre aux attentes des utilisateurs, qui ne peut, évidemment, se faire sans une étude approfondie de leurs besoins.

II.4.1. Identification des acteurs:

L'analyse des besoins exprimés nous a permis d'identifier cinq (5) utilisateurs potentiels:

- **Administrateur:** C'est l'ingénieur du département, il a l'accès à tout le système, le droit de modifier tous les utilisateurs (login, mot de passe, ex...) ainsi que d'affecter les rapports aux utilisateurs, ...etc.
- **Chef de département:** Ce dernier peut voir tout ce qui concerne son département avec ses différentes spécialités, son on adjoint peut aussi accéder avec les mêmes privilèges.
- **Responsable de la spécialité:** Chaque spécialité au niveau d'un département a son responsable. Ce dernier n'a la possibilité de voir que les informations concernant les étudiants de la spécialité dont il est responsable.
- **Responsable de la scolarité.**

II.4.2. Listes des besoins:

A travers notre étude des besoins et nos recherches, nous avons pu recenser les différents indicateurs nécessaires pour une bonne prise de décision au sein du département informatique, Les plus importants d'entre eux, sont classés dans le tableau suivant :

Indicateur	Catégorie	Description
Nombre d'enseignants par grade	camembert	Dans ce camembert on peut distinguer le nombre des enseignants selon leurs grades dans un département.
Taux de réussite des étudiants par spécialité	Graphique à barres	Ce graphique affiche le pourcentage de réussite d'une spécialité durant un semestre, l'axe des x la spécialité, l'axe y le pourcentage de réussite.

Evolution du nombre d'étudiants par année de chaque spécialité	line	Ce diagramme en ligne montre l'évolution des effectifs des étudiants d'un département répartis sur les différentes années, sur l'axe y on trouve le nombre des étudiants, sur l'axe x On y trouve les différentes années universitaires.
Nombre d'étudiants par observation	camembert (polar area)	Ce camembert affiche le nombre d'étudiants du département selon leur observation d'admission ou non admission, durant l'année universitaire (admis, admis session 2, admis dettes, ajourné).
Nombre d'étudiants par spécialité	Graphique à barres	Ce graphe permet aux décideurs 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.
Etat d'avancement des modules par section	Graphique à barres	Ce graphe permet aux décideurs de visualiser l'état d'avancement des modules par section, sur l'axe y on trouve le taux d'avancement et sur l'axe x les différents modules.

Tableau II.2 : Tableau récapitulatif des différents besoins

Conclusion

L'étude des besoins est une étape plus que nécessaire dans un projet Data Warehouse. C'est, en effet, à partir de cette étude que se décidera la manière de construction de l'entrepôt de données et de son architecture. Les besoins étant recensés, la construction du Data Warehouse peut alors commencer. Cette construction fera l'objet du chapitre suivant.

Chapitre III:

Conception de la solution

Introduction

Une fois les besoins des utilisateurs connus, nous pouvons commencer à concevoir notre Data Warehouse. Pour cela, nous avons eu recours à la modélisation dimensionnelle qui est souvent associée aux entrepôts de données compte tenu de ses avantages.

II.1. La base de données relationnelle:

Dans le chapitre précédent nous avons étudié la base de données opérationnelle du département mais il nous manquait les différentes relations entre les tables. Afin d'entamer la modélisation dimensionnelle de manière propre, nous avons jugé nécessaire de travailler à partir d'un schéma relationnel montrant les différentes relations¹.

II.2. La conception du datawarehouse:

II.2.1 Justification de notre approche

Les deux méthodes les plus connues dans la conception d'un entrepôt de données sont les approches introduites par Bill Inmon « top-down » et Ralph Kimball ¹« bottom-up ».

Ce qui les différencie peut être décrit comme la distinction entre l'extensibilité et la puissance de l'approche «top-down» contre la rapidité et la simplicité de l'approche « bottom-up ».

En vue des activités et services dont le département s'occupe, les Datamarts doivent être fonctionnels le plus tôt possible, d'où le choix de l'approche «Bottom-up» de Kimball qui consiste à commencer par concevoir les datamarts et les regrouper par la suite, jusqu'à l'obtention du DataWarehouse (DW).

II.2.2. La modélisation dimensionnelle:

Pour réaliser notre solution, nous avons opté pour le modèle en étoile jugé comme le modèle le plus adapté à notre application car il vise à optimiser le temps de réponse de requêtes complexes en agrégats.

Nous avons aussi réduit la hiérarchie des tables de dimensions, les entités de niveau supérieur peuvent être «réduites» en entités de niveau inférieur au sein des hiérarchies.

Dans le cas étudié, on a deux modèles de datamarts :

- Le datamart qui concerne L'enseignement.
- Le datamart qui concerne les statistiques concernant les étudiants.

Plus de détails sur la création des table fait/dimension sont données dans Annexe1.

¹voir Annexe3

II.2.2.1. Datamart « StudyEnseignement » :

Grain: Etat d'avancement d'un module dans chaque section, dans une certaine spécialité, un certain semestre, une certaine année.

Dimension:

- Dimension Temps: on calcule l'évolution du nombre par année.
- Dimension Affectation: Définit le pourcentage d'avancement de chaque module.
- Dimension Section: Définit la section de chaque spécialité de formation prise en charge par chaque enseignant.

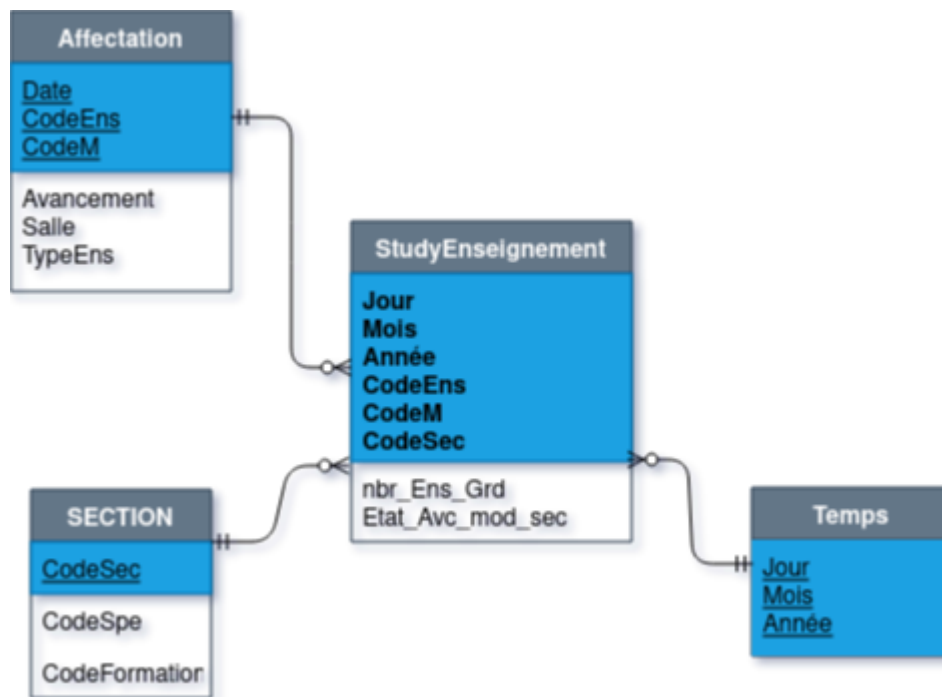


Figure III.1: Schéma en étoile "StudyEnseignement"

Détail de la table fait "StudyEnseignement":

Attribut	Description
codeEns	Code de l'enseignant.
codeM	code du module.
CodeSec	Code de section
Jour	jour de la date
Mois	mois de la date

Année	année de la date
Etat_Avan_Mod_Sec	Etat d'avancement des modules dans chaque section d'une spécialité.
Nbr_Ens_Grd	Nombre d'enseignants par grade.

Tableau III.1 : Tableau des attributs de la table fait "StudyEnseignement"

II.2.2.2. Datamart « StudyEtudiant » :

Grain: Suivie de toutes les statistiques concernant les étudiants: taux de réussite, nombre d'étudiants par observation, nombre d'étudiants par spécialité, évolution du nombre d'étudiant de chaque spécialité dans une certaine section, un certain semestre, une certaine année.

Dimension:

- Dimension Temps: L'évaluation du nombre et pourcentage par année.
- Dimension MoyenneS: Définit la moyenne semestrielle des étudiants.
- Dimension MoyenneM: Définit la moyenne modulaire des étudiants.
- Dimension MoyenneAnnu: Définit la moyenne annuelle des étudiants.

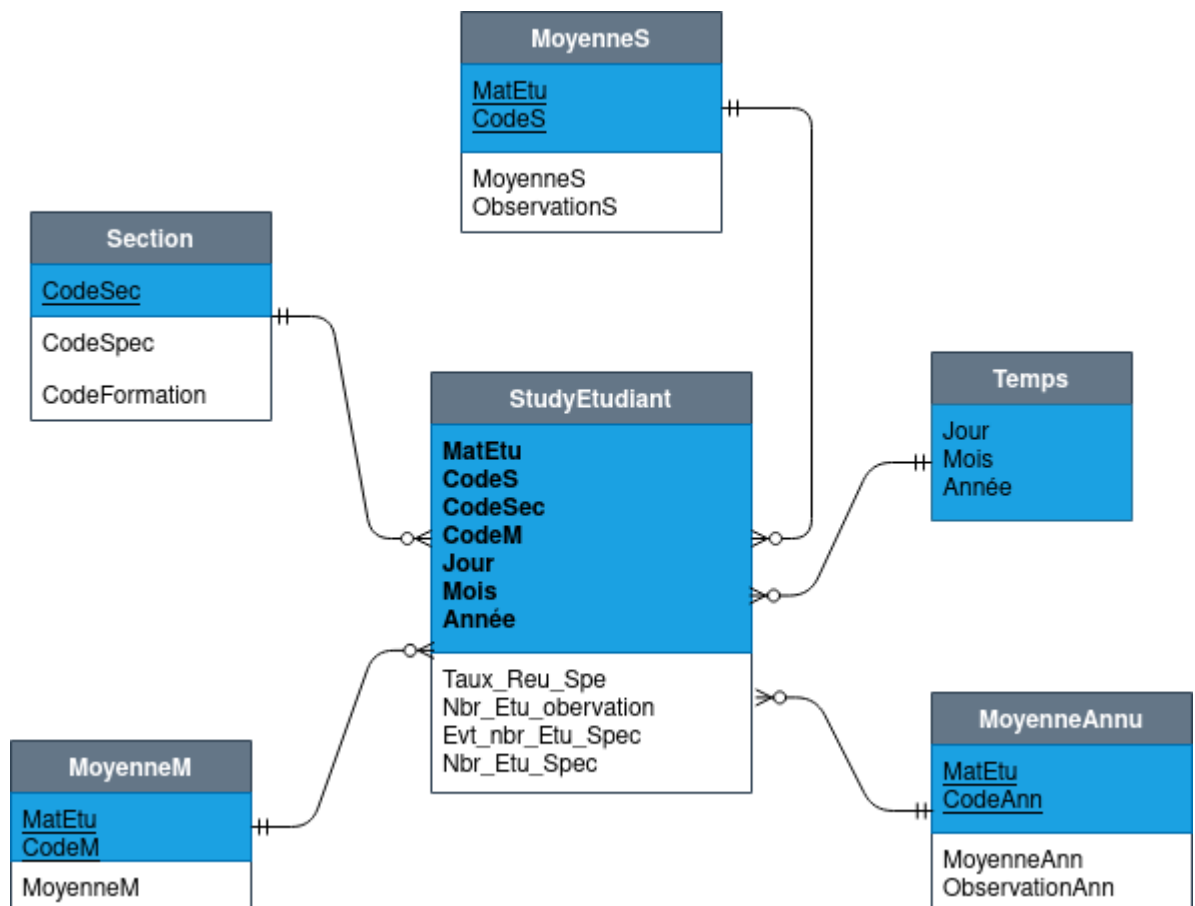


Figure III.2: Schéma en étoile "StudyEtudiant"

Détail de la table fait “StudyEtudiant”:

Attribut	Description
matEtu	matricule et clé étrangère de la dimension Etudiant
codeSpec	Code spécialité
codeSec	Code de section
codeM	code de module
Jour	jour de la date
Mois	mois de la date
Année	année de la date
Evt_nbr_Etu_Spec	Évolution du nombre d'étudiants pour chaque spécialité.
Nbr_Etu_Spec	Nombre d' étudiants par spécialité.
Taux_res_Spec	Taux de réussite des étudiants par spécialité.
Nbr_Etu_Obs	Nombre d' étudiants par observation.

Tableau III.2 : Tableau des attributs de la table fait “StudyEtudiant”

II.3. Alimentation du Data Warehouse

L'intégration des données définit le processus d'alimentation du DW.

II.3.1. L'extraction des données

L'extraction des données s'est faite à travers des ressources externes dans lequel nous avons essayé de simuler des bases de données opérationnelles qui ressemble à celles du département informatique afin de les utiliser ultérieurement dans l'entrepôt et ce par manque de temps.

II.3.2. Le chargement des données

Le diagramme d'activités suivant décrit le processus général de l'alimentation de l'entrepôt de données dès sa mise en service :

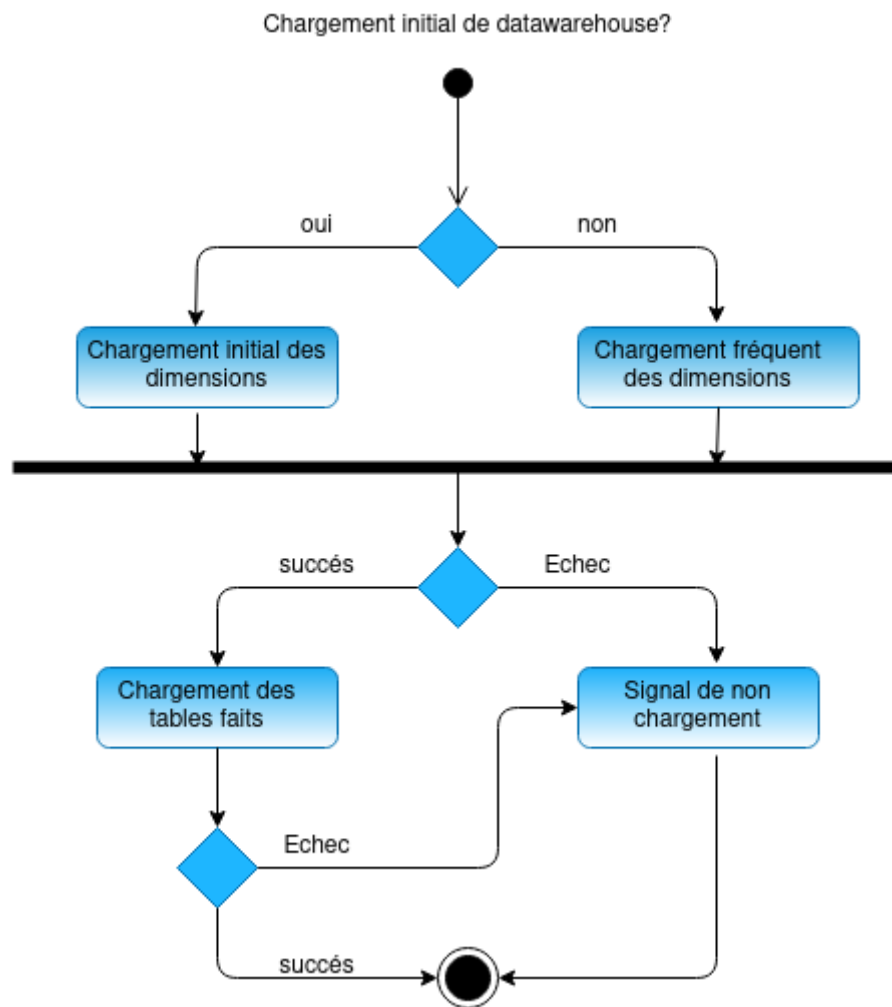


Figure III.3: Diagramme d'activité du processus de chargement du datawarehouse

Ce diagramme explique la phase de chargement des données dans la datawarehouse, On distingue deux types de tables dans l'entrepôt de données « faits, dimensions » Chaque type a des informations différentes. Le diagramme vise à comparer des chargement successif afin de détecter les différents changements dans l'entrepôt de données.

II.4. Diagramme des cas d'utilisation d'utilisation:

A travers ce diagramme nous allons visualiser les fonctionnalités globales de notre application.

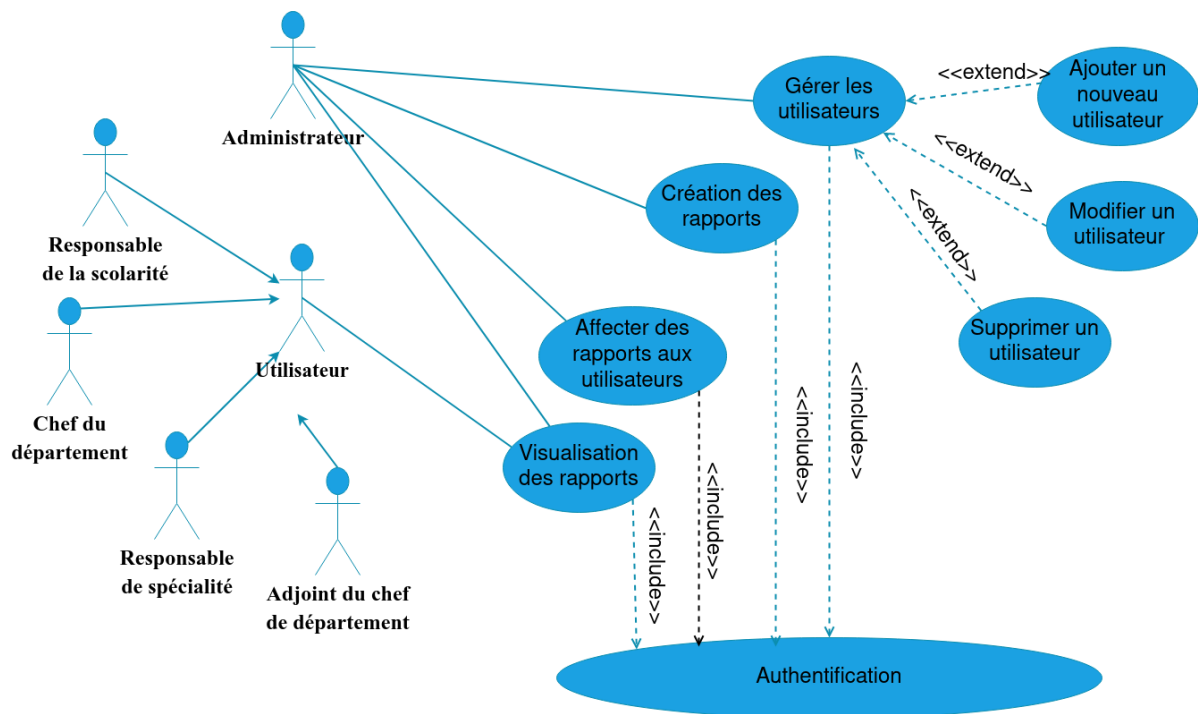


Figure III.4 : Diagramme des cas d'utilisation de l'application

Conclusion

La modélisation dimensionnelle suivie du processus ETL, permet d'avoir un datawarehouse exploitable permettant de naviguer et de manipuler les données, sans difficulté afin de satisfaire leurs besoins en analyse.

Le prochain et dernier chapitre va faire objet des différents tests et réalisation de notre datawarehouse.

Chapitre VI:

Réalisation et test

Introduction

Après avoir établi la conception de notre datawarehouse et exposer l'aspect théorique de notre étude, nous allons dans ce dernier chapitre, visualiser notre modélisation à travers un dashboard sur un site web, tout en expliquant les outils nécessaires utilisés pour la création de notre application.

VI.1. Environnement de travail

VI.1.1 Outils technologiques utilisés

PostgreSQL

PostgreSQL est un système de gestion de base de données relationnelle et objet, puissant et open source, connu pour ses performances par rapport aux bases de données volumineuses intégrant un ensemble d'outils d'administration et de configuration. Aussi ce SGBD est pré configuré pour la mise en place d'un Data Warehouse.[8]



Visual studio code:

Visual Studio Code (VSC) est un éditeur de code open-source, gratuit et multiplateforme (Windows, Mac et Linux), développé par Microsoft. Principalement conçu pour le développement d'applications avec JavaScript ,TypeScript et Node.js.[9]



Node js:

C'est un environnement d'exécution JavaScript orientée événements, open sources et multiplateforme, utilisant le moteur d'exécution V8 de Google chrome qui analyse et exécute du code JavaScript très rapidement et il fonctionne d'une manière asynchrone.

Express.js:

C'est un Framework d'application web le plus populaire pour node.js. Il rend le développement des applications web plus rapide et plus facile. Il offre aussi de nombreux outils pour rendre le développement plus optimisé.



Vue js:

C'est un framework évolutif pour construire des interfaces utilisateur. Il a été conçu et pensé pour pouvoir être adopté de manière incrémentale.



Chartjs:

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



VI.1.2. Langages utilisés

JavaScript

« JS », connu comme le langage de script des pages web. Il est aussi utilisé dans de nombreux environnements extérieurs aux navigateurs web tels que Node.js, Apache.[10]



HTML5:

C'est le langage de balisage pour la structure et la présentation des pages Web dans le navigateur, il est souvent utilisé avec le JavaScript et le CSS.



CSS3:

C'est un langage de style utilisé pour décrire comment les éléments HTML doivent être affichés sur l'écran, couleur, police, bordures,..etc.



SQL:

C'est un langage de requête structurée qui sert à effectuer des opérations sur des bases de données. Parmi ces opérations nous pouvons citer : l'ajout, la modification et la suppression de données dans la base de données, et c'est le langage utilisé par les SGBD tel que Oracle, MySql, Postgresql. . . etc.

VI.2. Présentation de l'application

Nous allons dans cette partie faire la présentation de notre application

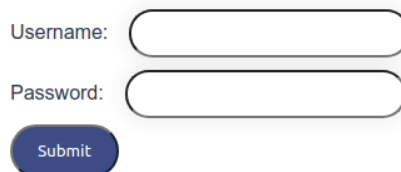
❖ Page d'accueil

Dès que l'utilisateur ouvre le site, la page suivante s'ouvrira.



Figure VI.1 : Page d'accueil du site

❖ Page de connexion



Username:

Password:

Figure VI.2 : Page d'authentification

❖ Interface Administrateur:

L'ingénieur a la possibilité d'ajouter, supprimer, modifier les utilisateurs, ajouter d'autres indicateurs, affecter les indicateurs aux utilisateurs et enfin visualiser les différents indicateurs.

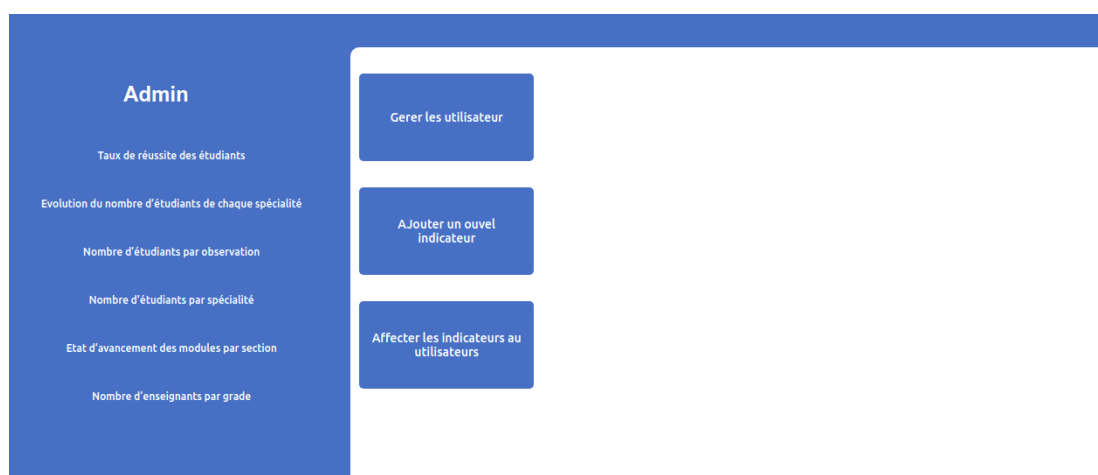


Figure VI.3 : Interface Administrateur

❖ Interface Utilisateur:

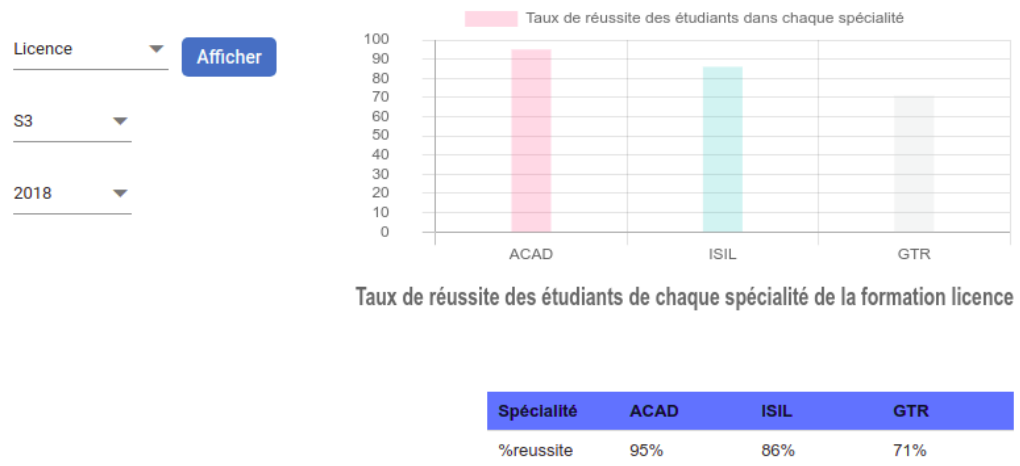
Le chef de département, son adjoint ou le responsable de scolarité ont la possibilité de visualiser tous les indicateurs. Le responsable de spécialité a la possibilité de visualiser que les indicateurs liés à sa spécialité.



Figure VI.3 : Interface des indicateurs

❖ Interface de consultation des indicateurs:

Une fois l'utilisateur authentifié, il accède à une interface, où il peut consulter les différents indicateurs listés dans l'analyse des besoins.



Taux de réussite des étudiants de chaque spécialité de la formation licence

Figure VI.4 : Interface de l'indicateur "Taux de réussite des étudiants"

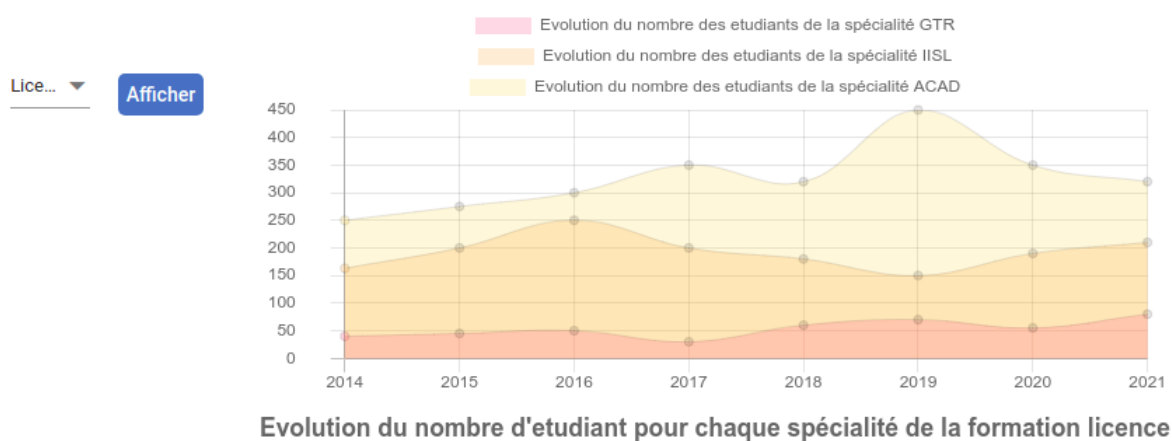


Figure VI.5 : Interface de l'indicateur "Evolution du nombre d'étudiants dans chaque spécialité"

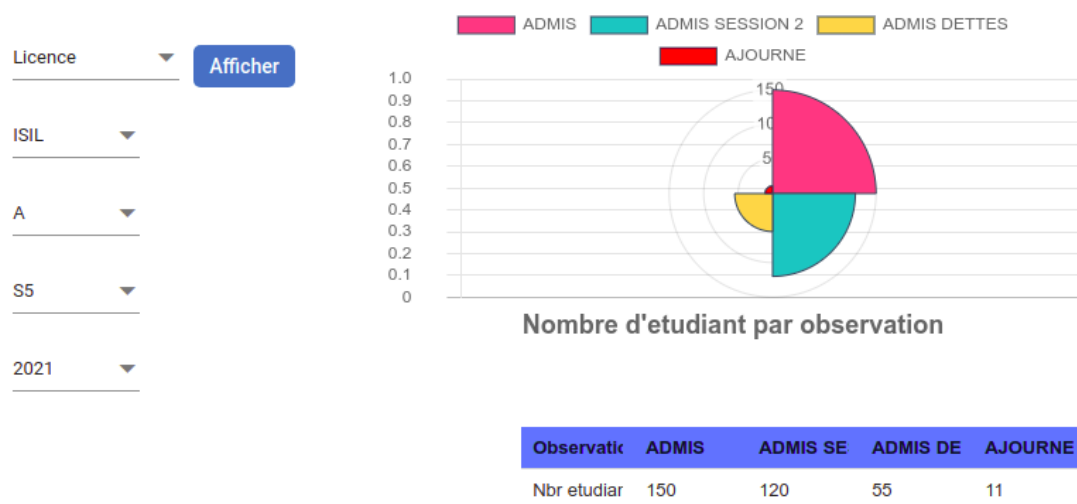
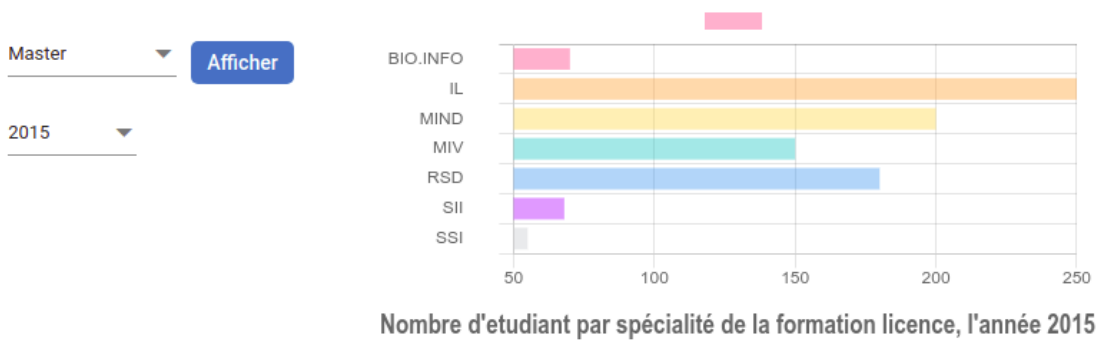
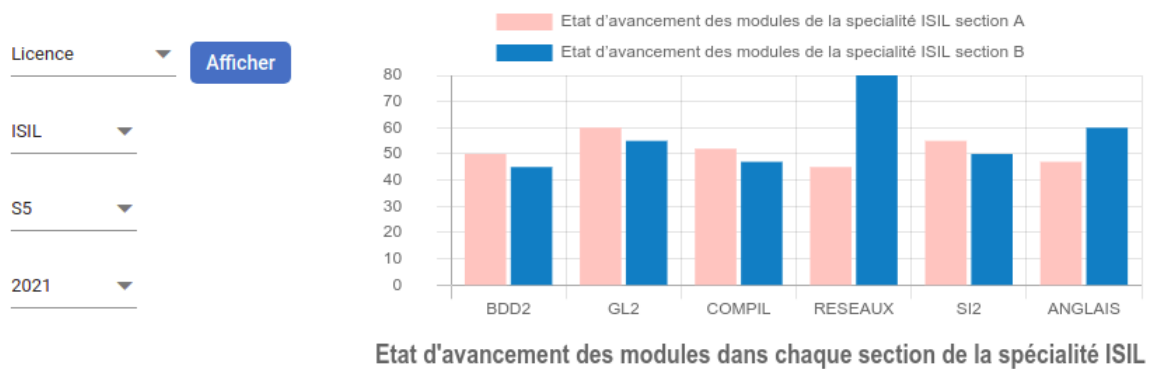


Figure VI.6 : Interface de l'indicateur "Nombre d'étudiants par observation"



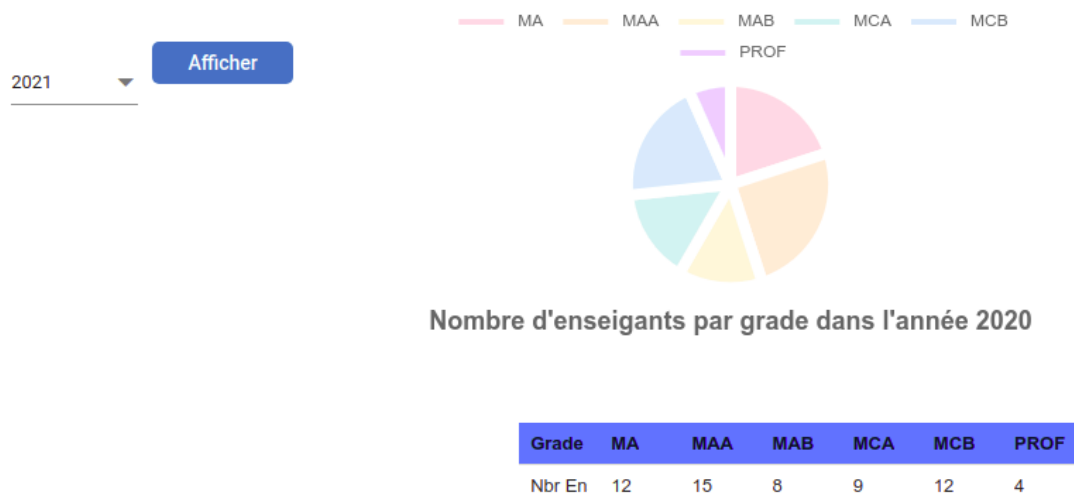
Spécialité	Nbr étudiant
IL	250
BIO.INFO	70
SSI	55
SII	68
RSD	180
MIND	200
MIV	150

**Figure VI.6 : Interface de l'indicateur
"Nombre d'étudiant par spécialité"**



Module	BDD2	GL2	COMPIL	RESEAUX	SI2	ANGLAIS
Section A	50%	60%	52%	45%	55%	47%
Section B	45%	55%	47%	80%	50%	60%

**Figure VI.7 : Interface de l'indicateur
"Etat d'avancement des modules par section"**



**Figure VI.8 : Interface de l'indicateur
"Nombre d'enseignant par grade"**

Conclusion

Dans ce chapitre, nous avons présenté l'environnement de travail dans lequel on avait implémenté notre application, cette dernière a été mise en place avec des outils «Open Source » que nous avons décrits dans ce chapitre, et pour finir nous avons donné un aperçu sur notre application en exposant les interfaces graphiques qui la composent avec une description de chacune des captures d'écran.

Conclusion Générale

Conclusion générale et Perspectives

Les entreprises d'aujourd'hui ont de plus en plus du mal à gérer les ressources propres à leur établissement, et cela est dû à la quantité massive des données et la façon dont elle est exploitée.

Dans le cadre de notre projet, et afin de pallier aux problèmes récurrents dans le processus de prise de décision au niveau du département informatique, on nous a initié le projet de réalisation d'un Data Warehouse pour gérer l'activité de notre établissement.

Nous avons exploré en premier lieu, le concept en général, ce qui nous a permis d'avoir une idée globale sur les différentes méthodes existantes pour la conception et la mise en œuvre d'une datawarehouse.

Nous avons ensuite enquêté sur le département informatique, sa structure, ses services ainsi que les outils utilisés pour sa gestion. Nous avons profité aussi de cette phase pour discuter sur les éventuels besoins qui peuvent intéresser certains utilisateurs au niveau du département.

Nous avons entamé après, la modélisation dimensionnelle de notre application et ce en suivant un modèle en étoile.

Et enfin pour visualiser les résultats obtenus de notre analyse des besoins et de l'existant, nous avons travaillé sur un dashboard et ce en utilisant des outils open source.

Ce projet nous a été d'une très bonne et productif expérience et un facteur important qui nous a enrichi avec de nouveau concepts lié à la BI, data science...etc

Nous avons éventuellement touché à la création des sites web, ou on a appris le langage javascript et l'utilisation des frameworks express et vuejs.

Des perspectives et améliorations peuvent être envisagées pour ce travail, citons principalement :

- Alimentation de la datawarehouse avec les tables opérationnelles du département.
- Étendre le déploiement de manière à couvrir tous les services de scolarité de l'université.
- Continuer le développement d'autres indicateurs non programmés.
- Intégrer plus de fonctionnalités pour améliorer le système.
- Rendre le site plus réactif à son utilisateur.

Bibliographie

Ouvrages

- [Dresner, 2001] : H. Dresner ; « BI : Making the Data Make Sense » ; Gartner Group 2001.
[Inmon, 2002]: W. H. Inmon ; «Building the Data Warehouse».
[Kimball, 2002] : R. Kimball et M. Ross ; Entrepôts de Données : Guide Pratique de Modélisation Dimensionnelle 2ème édition » ; Vuibert 2002.

Articles et Thèses

- [1]: Synthétisation à partir de la thèse de Ronan Tournier « Analyse en ligne (OLAP) de documents» [Ronan Tournier, 2009].
[2]: Article University of Technology of Compiègne «Data warehouse et outils décisionnels»[STÉPHANE CROZAT, 15 septembre 2016].
[3]: Document «From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design ,p5» [Daniel L. Moody, 2003]
[4]: Mémoire de master “Conception et réalisation d’un Data Warehouse pour la mise en place d’un système décisionnel”[2009/2010]
[6]: Mémoire de licence Option Informatique Générale: Thème Conception d’un tableau de bord scolarité licence[2016]
[7]: Plan de Développement Informatique de l'USTHB pour la période 2010 – 2014

Webographie

- [5]: <http://deptinfo.usthb.dz>
[8]: <https://www.oracle.com/fr/database/definition-postgresql.html>
[9]: <https://edutechwiki.unige.ch>
[10]: <https://developer.mozilla.org/fr/docs/Web/JavaScript>

Listes des figures

Figure 1: Diagramme de Gantt de l'avancement du projet.....	3
Figure I.1: Le système d'aide à la décision.....	4
Figure I.2: Le décisionnel au sein du Système d'information [Goglin, 1998].....	6
Figure I.3: évolution des bases de données décisionnelles.....	8
Figure I.4: The structure of the data warehouse.....	8
Figure I.5: Architecture globale d'un Data Warehouse.....	9
Figure II.1: Organigramme du département informatique.....	11
Figure III.1: Schéma en étoile "StudyEnseignement".....	16
Figure III.2: Schéma en étoile "StudyEtudiant".....	17
Figure III.3: Diagramme d'activité du processus de chargement du datawarehouse.....	19
Figure III.4: Diagramme des cas d'utilisation de l'application.....	20
Figure VI.1: Page d'accueil du site.....	22
Figure VI.2: Page d'authentification.....	23
Figure VI.3: Interface Administrateur.....	23
Figure VI.4: Interface de l'indicateur "Taux de réussite des étudiants".....	24
Figure VI.5: Interface de l'indicateur "Evolution du nombre d'étudiants dans chaque spécialité".....	25
Figure VI.6: Interface de l'indicateur "Nombre d'étudiants par observation".....	25
Figure VI.6: Interface de l'indicateur "Nombre d'étudiant par spécialité".....	26
Figure VI.7: Interface de l'indicateur "Etat d'avancement des modules par section".....	26
Figure VI.8: Interface de l'indicateur "Nombre d'enseignant par grade".....	27

Liste des tableaux

Tableau I.1 : Tableau comparatif entre les tables de faits et les tables de dimensions.....	7
Tableau II.1 : Tableau des données de la table etudiant.....	14
Tableau II.2 : Tableau récapitulatif des différents besoins.....	17
Tableau III.1 : Tableau des attributs de la table fait "StudyEnseignement".....	17
Tableau III.2 : Tableau des attributs de la table fait "StudyEtudiant".....	18

Annexe

Annexe1:

Les détails de la conception et la modélisation dimensionnelle d'un entrepôt de données

Dans cette partie on décrit une méthode pour le développement d'un modèle dimensionnel à partir d'un modèle opérationnel.

Etape 01: Classification des entités

La première étape de la modélisation dimensionnelle est la classification des entités en trois modèles:

- **Entités transactionnelles:**

Les entités transactionnelles enregistrent des détails sur des éléments particuliers des événements qui se produisent dans une entreprise à un moment donné et contiennent des mesures ou des quantités, par exemple: les commandes, les réclamations d'assurance, les paiements de salaire et les réservations d'hôtel. Invariablement, ce sont ces événements que les décideurs veulent pour comprendre et analyser.

Les entités transactionnelles sont les entités les plus importantes dans un entrepôt de données et constituent la base pour construire **des tables de faits** dans des schémas en étoile.

- **Entités constitutives :**

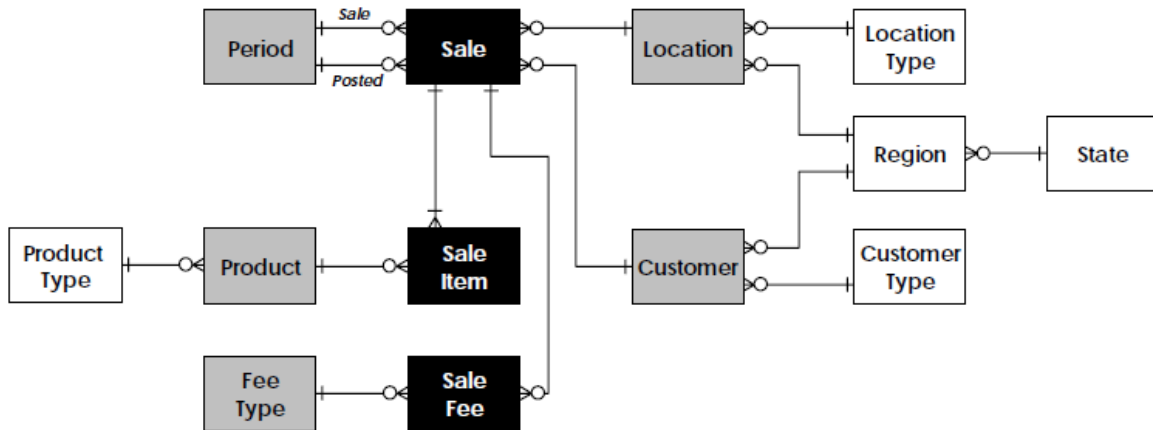
Une entité constitutive (composante) est une entité qui est directement liée à une entité transactionnelle via une relation un-à-plusieurs. Les entités composantes répondent aux questions « qui », « quoi », « quand », « où », « comment » et « pourquoi » d'un événement d'affaires.

- **Entités de classification:**

Les entités de classification sont des entités liées aux entités constitutives par une chaîne de relations individuelles, c'est-à-dire qu'elles dépendent fonctionnellement d'une entité composante (directement ou transitivement). Les entités de classification représentent **les hiérarchies intégrées** dans les modèles de données, qui peuvent être regroupées en entités composantes pour former des **tables de dimensions** dans un schéma en étoile. Dans certains cas, les entités peuvent entrer dans plusieurs catégories.

Nous définissons donc une hiérarchie de préséance pour résoudre de telles ambiguïtés :

1. Entité transactionnelle (priorité la plus élevée)
2. Entité de classification
3. Entité composant (priorité la plus basse)



Classification des entités

- Les entités noires représentent les entités de transaction.
- Les entités grises indiquent les entités de composant.
- Les entités blanches indiquent les entités de classification.

Etape 02: Identification de la structuration hiérarchique (schéma d'étoiles)

Chaque dimension d'un schéma d'étoiles consiste naturellement en une ou plusieurs hiérarchies intégrées dans n'importe quelle séquence des entités réunies par des relations individuelles, tous alignés dans la même direction. Ceux-ci fournissent une façon de classer les événements d'entreprise stockés dans la table de faits, diminuant ainsi la complexité.

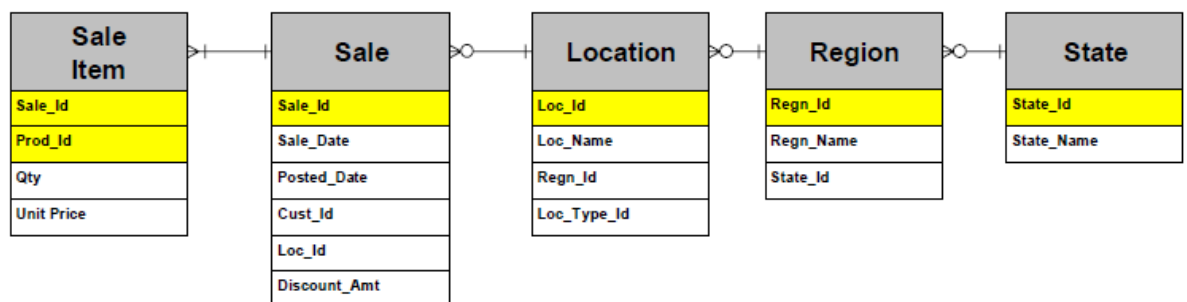


Figure I.9 : Exemple de hiérarchie

Etape 03: Conception détaillée du table des faits

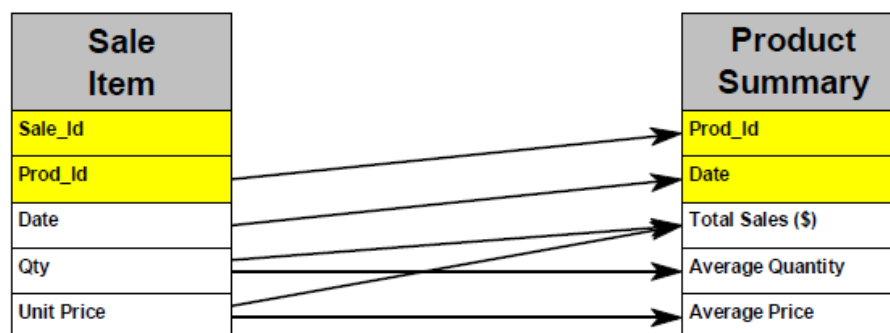
a). Identification de la clé primaire du table des faits :

La clé d'une table de faits comprend toutes les clés dimensionnelles.

b). Identification des faits:

Les attributs non clés des tables des faits sont des faits qui peuvent être évalués à l'aide de fonctions numériques. Ces attributs sont stockés dans des entités transactionnelles.

On peut appliquer **des opérations d'agrégation** à une entité transactionnelle pour obtenir une table des faits contenant un résumé des données. Un sous-ensemble d'attributs est choisi parmi les entités sources à agréger (attributs d'agrégation) et un autre sous-ensemble d'attributs choisis pour être agrégés par (les attributs de regroupement). Les attributs d'agrégation doivent être des quantités numériques.



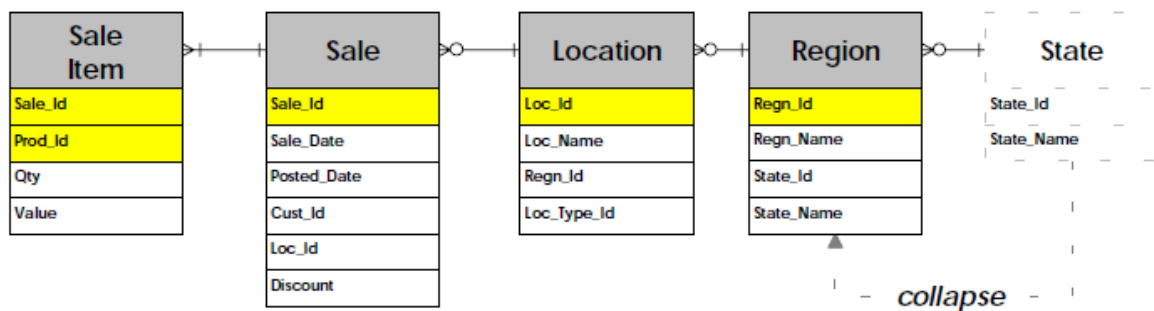
Etape 04: Conception détaillée des tables des dimensions

a). Identification de la clé primaire du chaque table de dimension:

La clé de chaque table de dimensions doit être un attribut numérique unique. Alors il faut généraliser la clé opérationnelle pour s'assurer qu'elle reste unique au fil du temps.

b). Réduction des hiérarchies:

Les tables de dimensions sont construites par l'effondrement ou la dénormalisation des hiérarchies en entités constitutives alors les entités de niveau supérieur peuvent être «réduites» en entités de niveau inférieur au sein des hiérarchies.



Annexe2:

Tableau des données de la table moyenneAnn:

Nom de données	Code de donnée	Type de donnée	Taille de donnée
moyenne annuelle	moyenneAnn	réel	
observation annuelle	observationAnn	chaîne	50

Tableau des données de la table CycleAnnu

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code annuel	codeAnn	alphanumérique	2

Tableau des données de la table Semester

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de semestre	codeSec	alphanumérique	2

Tableau des données de la table MoyenneSem

Nom de données	Code de donnée	Type de donnée	Taille de donnée
moyenne de semestre	moyenneS	réel	
observation	observationS	chaîne	50

Tableau des données de la table Groupe

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de groupe	codeGrp	alphanumérique	2
libele de groupe	libeleGrpc	chaîne	10

Tableau des données de la table Section

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de section	codeSec	alphanumérique	2
libele de section	libelle	chaîne	10

Tableau des données de la table Spécialité

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de spécialité	codeSec	alphanumérique	10
libele de spécialité	libelle	chaîne	10
disignation de spécialité	disignationSpec	chaîne	20

Tableau des données de la table Formation

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de formation	codeForm	alphanumérique	10
libele de formation	libelleForm	chaîne	10
disignation de formation	disignationFor	chaîne	20

Tableau des données de la table Enseignant

Nom de données	Code de donnée	Type de donnée	Taille de donnée
numéro d'enseignant	numEns	numérique	12
nom	nomEns	chaîne	50
prenom	prenomEns	chaîne	50
email	emailEns	chaîne	50
téléphone	NTphEns	numérique	10

grade	gradeEns	chaîne	250
nombre d'heures d'enseignement	nb_heurs_Ens	numérique	3

Tableau des données de la table Module

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de module	codeM	alphanumérique	30
coefficient de module	coef	numérique	2
credit	creditM	numérique	2
note de module	noteM	réel	

Tableau des données de la table MoyenneMod

Nom de données	Code de donnée	Type de donnée	Taille de donnée
moyenne	codeM	réel	
observation	observationM	chaîne	20

Tableau des données de la table Affectation

Nom de données	Code de donnée	Type de donnée	Taille de donnée
date	date	numérique	8
type	typeEns	chaîne	20
salle	creditM	chaîne	10

Tableau des données de la table UnitéEns

Nom de données	Code de donnée	Type de donnée	Taille de donnée
codeUnité	date	numérique	10
type d'unité	typeEns	chaîne	20

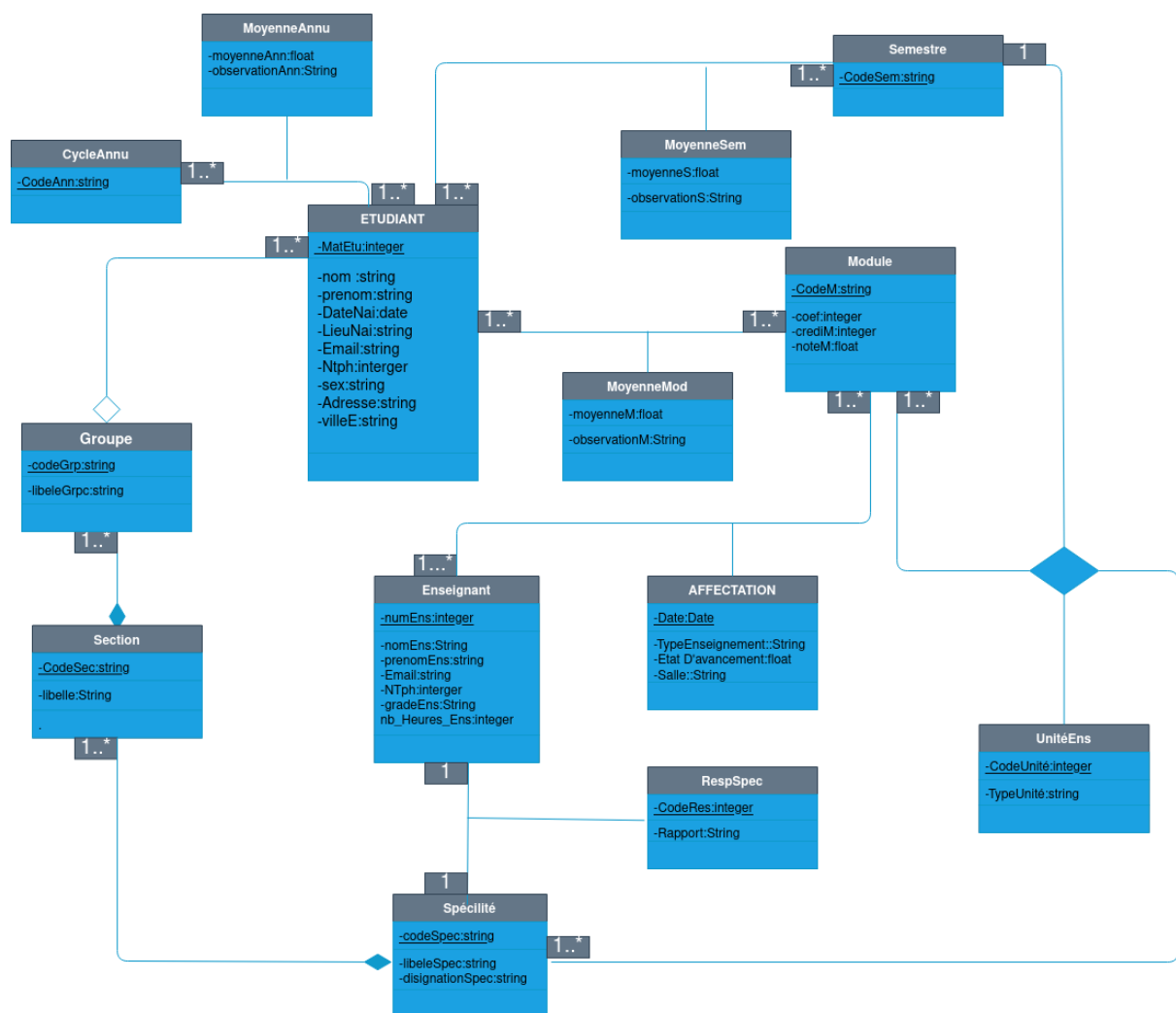
Tableau des données de la table RespSpec

Nom de données	Code de donnée	Type de donnée	Taille de donnée
code de res	codeRes	numérique	10
rapport	typeEns	chaîne	20

Annexe3:

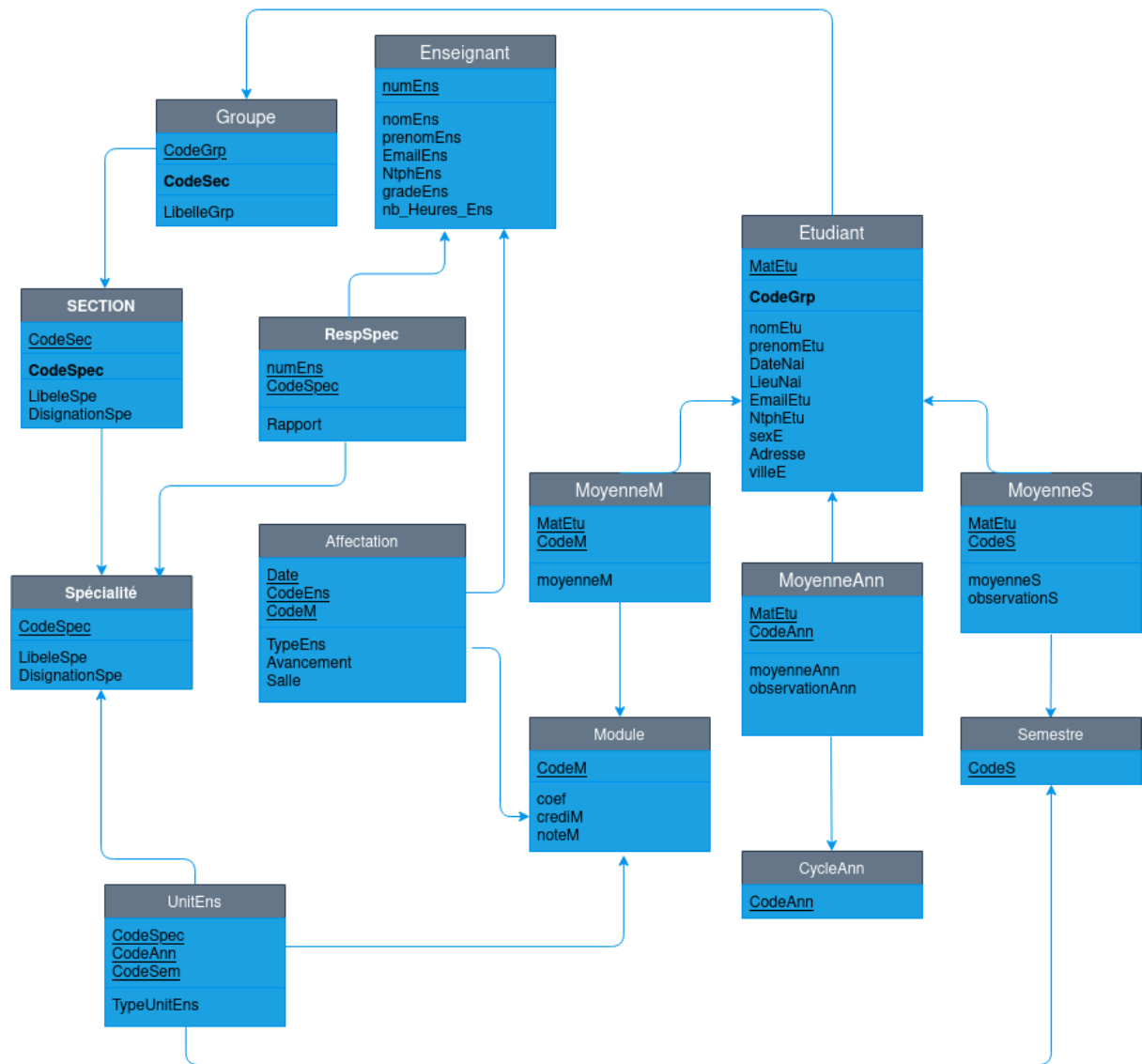
Le Diagramme de classe:

Nous avons élaboré un diagramme de classe pour en extraire ensuite le schéma relationnel de l'activité du département.



Le schéma relationnel:

Après l'analyse du diagramme de classe, nous pourrions visualiser les tables et leurs relations dans les schéma relationnelles ci-joint:



Résumé

Le département informatique de l'université des sciences et de la technologie Houari Boumediene, recueille chaque année une quantité considérable de données qui sont structurées de manière hétérogène. L'élaboration d'un système décisionnel est donc nécessaire pour homogénéiser ces données et mettre à la disposition de la direction du département Informatique un outil d'aide à la décision en s'appuyant sur l'analyse des indicateurs pertinents.

Notre étude vise à la conception et mise en place d'un système en mesure d'exploiter les données issues des bases opérationnelles du département, et ainsi offrir des informations de qualité aux décideurs. Ce système sera construit autour d'une base de données dédiée totalement aux décisionnel un « Data Warehouse » et répondant à tous les besoins d'analyse du groupe dans sa fonction de distribution. Ce présent projet a donc pour vocation première de réaliser une telle base de données.

Mots clés : Data Warehouse « Entrepôt de données », Système Décisionnel.