

UNE APPROCHE DE CONCEPTION DES OUTILS D'AIDE A LA DECISION BASES SUR DATAWAREHOUSE

Engombe Wedi Boniface¹, Okit'oleko On'okoko Wenga², and Kafunda Katalay³

¹Université Pédagogique Nationale, RD Congo

²Institut Supérieur de Statistique de Kinshasa, RD Congo

³Université de Kinshasa, RD Congo

Copyright © 2020 ISSR Journals. This is an open access article distributed under the *Creative Commons Attribution License*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT: The work carried out consists of a framework for evaluation and support of a project to develop decision support tools based on data warehouse for the management of the forecast of Higher Education Institutions and Universities, EESU/HEIU in acronym. Currently, these tools occupy a prominent position in the evolution perspective of the management of infrastructure capacity according to student enrolment for each academic year by 2090 using the simple linear regression line. In this article, we are working on designing, implementing, and securing a Meta Object – class data warehouse deployed in a high – speed, optical fiber backbone environment. Then, we were interested in realizing EESU/HEIU management software around the nine most popular targeted applications, thanks to the WINDEV development platform.

KEYWORDS: approach, design, support tools, decision, data warehouse.

RESUME: Le travail effectué consiste en un cadre d'évaluation et d'accompagnement d'un projet de développement des outils d'aide à la décision basés sur datawarehouse pour la gestion prévisionnelle des Etablissements d'Enseignement Supérieur et Universitaire, EESU en sigle. Actuellement, ces outils occupent une place prépondérante dans la perspective évolutive de la gestion prévisionnelle de la capacité d'accueil des infrastructures en fonction des effectifs des étudiants pour chaque année académique à l'horizon 2090 par usage de la droite de régression linéaire simple. Dans cet article, nous travaillons sur la conception, l'implémentation et la sécurisation d'un méta entrepôts objet-classe de données déployé dans un environnement backbone à fibre optique connecté sur Internet à haut débit. Ensuite, nous nous sommes intéressés à réaliser un logiciel de gestion des EESU autour des neuf applications les plus répandues ciblées grâce à la plateforme de développement WINDEV.

MOTS-CLEFS: approche, conception, outils d'aide, décision, datawarehouse.

1 INTRODUCTION

Cet article se veut être constitué un document de référence pouvant servir demain à la mise en place et au déploiement d'un système informatique via un backbone à fibre optique autour d'un Méta modèle Entrepôt objet-classe de données sécurisé, piloté par des outils d'aide à la décision en vue de la gestion rationnelle des ressources des Etablissements d'Enseignement Supérieur et Universitaire. Il s'agit d'un mode de gestion évoluant au rythme des nouvelles technologies de l'information et de communication, NTIC en sigle, des entrepôts de données complexes. Le projet poursuit les objectifs qui suivent :

- Etudier les approches de conception existantes en vue de proposer une approche optimale et rationnelle du type hybride pour pallier aux insuffisances des approches précédentes ;
- Réaliser, dans la pratique, un Méta Entrepôt de données orienté objet-classe sécurisé comprenant 9 modules des ressources logicielles intégrées notamment webmail, LMS, bureau numérique, gestion des ressources humaines et académiques, infrastructures, budget, matérielles et logicielles dont voici l'architecture globale ci-après ;
- Constituer un modèle de base d'un backbone à fibre optique haut débit de déploiement de Méta modèle Entrepôt des données pour un éventuel partage en ligne du type Internet ;
- Proposer un arbre de décision organisationnel hiérarchique de l'ESU avec perspective également usage de droite de régression susceptible de prédire la capacité d'accueil dans chaque Etablissement d'Enseignement Supérieur et Universitaire en fonction des effectifs croissants des étudiants enregistrés à l'horizon 2090 par exemple ;
- Renforcer la sécurité multi niveaux dans le backbone vers l'Internet haut débit pour lutter contre les pirates, les malveillants notamment la cybercriminalité, cyberattaque qui font l'objet d'actualité pour le moment à travers le monde entier.

Pour atteindre ces objectifs, il nous paraît nécessaire d'indiquer les différents points qui seront traités à savoir :

- Les modes interdisciplinaires des domaines de recherche ;
- Les généralités sur les systèmes décisionnels ;
- Un état de lieux du secteur éducatif des Etablissements d'Enseignement Supérieur et Universitaire, EESU en sigle ;
- La fouille de données ;
- Une approche de conception préconisée ;
- Implémentation de méta modèle entrepôt de données orienté objet-classe ;
- Développement du logiciel à Windev 20 ;
- Déploiement du logiciel dans un backbone à fibre optique en toute sécurité.

1.1 CHOIX ET INTERNET DU SUJET

Le ministère de l'Enseignement Supérieur et Universitaire en RD Congo a le souci de bâtir un système d'information stratégique d'aide à la prise des décisions pour la bonne gouvernance conformément à l'évolution des TIC.

Par le présent article, nous avons démontré comment il est possible d'exploiter les bases de données décisionnelles unifiées orientées métiers au sein des Etablissements d'Enseignement Supérieur et Universitaire afin de réaliser un méta modèle objet-classe d'Entrepôts de données de gestion de cette juridiction.

1.2 METHODOLOGIE

Notre première démarche a consisté à prendre respectivement l'apprentissage artificiel, une portion de l'intelligence artérielle, en interaction avec les modes interdisciplinaires analyse de données, automatique, la vision assistée par ordinateur et l'optimisation.

Secundo, la deuxième étape consiste à faire l'état des lieux des approches de conception des outils d'aide à la décision existantes en vue dégager une approche dite mixte comme une solution optimale un envisagée.

Tertio, nous avons présenté le logiciel développe sous WinDev et déployer en réseau à backbone à fibre optique avec l'usage de la droite de régression prédictive de dépendance effectifs étudiants et capacité d'accueil des infrastructures.

1.3 CONTRIBUTIONS

Nous avons présenté, au niveau de la problématique, les différents problèmes relatifs aux entrepôts de données complexes auxquels nous nous sommes intéressés pour proposer une solution optimale d'usage de l'approche de conception hybride des outils d'aide à la décision basés sur DataWarehouse pour la gestion intégrée des ressources des EESU en RDC plus particulièrement la ville de Kinshasa.

Ainsi, les problèmes abordés constituent effectivement les différents verrous scientifiques du système décisionnel du 21ème auxquels nous avons tenté d'apporter personnellement des solutions optimales adaptées à l'environnement ciblé

conformément à l'évolution des bases de données complexes décisionnelles constituant un Méta modèle entrepôt de données objet-classe.

A cet effet, le point focal de nos travaux est la prise en compte du décideur des EESU dans le processus d'entreposage et de partage de données très volumineuses orientées sujets et ou services.

L'originalité de nos travaux de recherche a consisté à concevoir et développer un Méta modèle entrepôt de données pour la gestion intégrée des ressources des EESU dans la ville province de Kinshasa tout en intégrant les techniques de la fouille de données en ligne au sein d'un SGBD Oracle pour traiter des grandes bases et ou entrepôts de données sans limitation de taille de stockage en fournissant des axes d'analyse pertinents dans des intervalles des temps très réduits de l'ordre des quelques secondes.

Ensuite, nous avons montré l'avantage de combiner de la fouille de données avec le Méta modèle entrepôt de données objet-classe en analyse multidimensionnelle pour la création des nouveaux axes d'analyse sémantique plus riches grâce à l'opérateur d'agrégation sémantique du type Roll-Up With K-Means OKITO, permettant de créer un nouveau niveau de granularité dans une hiérarchie de dimension spécifique. K- Means est une méthode de classification automatique qui consiste d'extraire des nouvelles structures sémantiques plus riches pour poser des nouvelles requêtes décisionnelles sur le Méta entrepôt de données regroupant neuf entrepôts des ressources de gestion des EESU. Puis le déployer en toute sécurité en réseau intersites du type backbone à fibre optique avec perspective de connexion à Internet à haut débit

Enfin, en perspective, expérimenter en statistique la droite de régression pour estimer les effectifs des étudiants de chaque année académique pour la construction des infrastructures au-delà de l'horizon 2090 par exemple.

1.4 SUBDIVISION DU TRAVAIL

Pour atteindre les objectifs visés, nous avons organisé ce travail en :

- Les modes interdisciplinaires des domaines de recherche ;
- Les généralités sur les systèmes décisionnels ;
- L'état de l'art des approches de conception des outils d'aide à la décision ;
- Fouille de données en ligne via OLAP ;
- Une approche de conception des outils d'aide à la décision bases sur Datawarehouse ;
- Implémentation et déploiement sécurisé du Meta Modèle Entrepôts objet-classe EESU conçu.

2 LES MODES INTERDISCIPLINAIRES DES DOMAINES DE RECHERCHE

2.1 BUT

Cette étape est de présenter les différents domaines cibles interagissant avec l'intelligence artificielle dans notre travail de recherche notamment l'analyse de données en statistiques, l'automatique avec les indicateurs de performance et les régulateurs correcteurs PID, l'optimisation par usage de l'approche LMI et la technique Tuning, l'apprentissage artificiel et afin la vision bayésienne.

2.2 SCHÉMA FONCTIONNEL

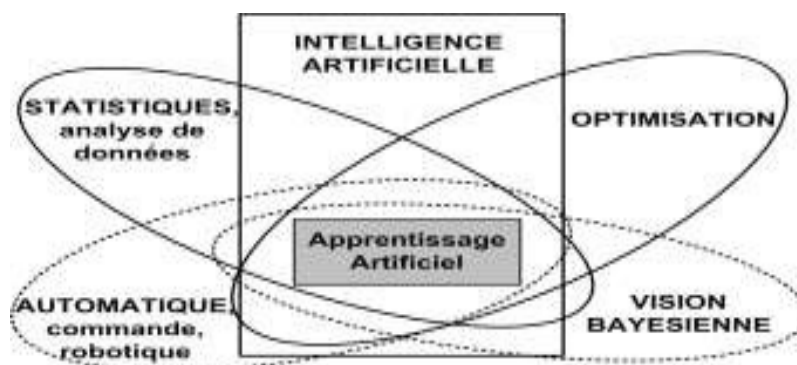


Fig. 1. Modes interdisciplinaires de recherche [3]

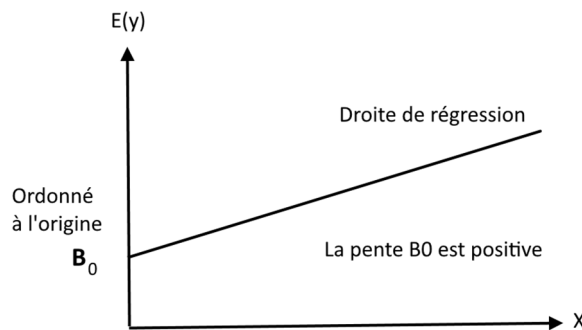
2.3 ETUDE DE L'ESPACE DE RECHERCHE PAR DROITE DE REGRESSION LINEAIRE

Elle consiste d'établir la dépendance fonctionnelle entre deux variables x et y pour l'apprentissage prédictif de la gestion prévisionnelle des infrastructures en fonction des effectifs des étudiants.

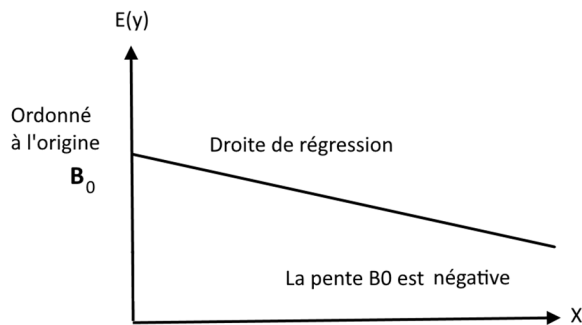
2.3.1 PRÉSENTATION [4]

La droite de régression linéaire est une droite qui décrit la relation entre la variable dépendante et la variable indépendante simple dite explicative. La variable dépendante est désormais notée X. La relation Y et X est supposée n'être qu'approximative et elle est perturbée par un terme d'erreur additif notée ϵ où par hypothèse $E(\epsilon) = 0$. La formule est donnée par : $Y = \beta_0 + \beta_1 X + \epsilon$ avec $\epsilon = 0$ ainsi la formule devient : $Y = \beta_0 + \beta_1 X$ où β_0 et β_1 sont les paramètres du modèle mais le terme d'erreur ϵ est une variable aléatoire. Donc β_0 représente le point d'intersection de présente la pente de la droite de régression. Elle donne le nombre d'unités supplémentaires de Y associées à une augmentation par unité de X. la droite de régression avec l'ordonnée dite intercepter constante et β_1 . $E(Y/X)$ est la moyenne de Y pour une valeur de X donnée unité de Ainsi, trois cas possibles peuvent exister de la manière suivante :

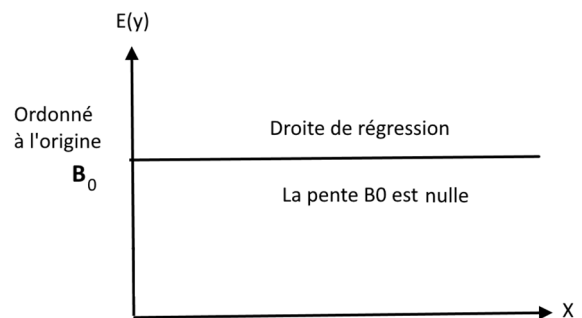
Premier cas : la relation linéaire positive



Deuxième cas : la relation linéaire négative



Troisième cas : la relation n'existe pas



L'équation estimée :

Elle a pour but d'estimer les valeurs des paramètres de la population inconnus θ_0 et θ_1 qui sont des coefficients de régression où b_0 et b_1 servent d'estimations de θ_0 et θ_1 d'où l'équation estimée de la régression est donnée par : $Y = \theta_0 + \theta_1 x$ avec y représente l'estimation ponctuelle de $E(y/x)$. Cependant, une équation estimée de la régression n'établit pas la relation de cause à effet entre x et y . Elle indique seulement comment ou dans quelle mesure les variables sont associées ou la corrélation différente de la causalité.

2.3.2 BUT POURSUIVI

Il s'agit tout simplement de déterminer un outil d'ajustement affine prédictif représentant une série statistique ou un nuage de points de la forme (x_i, y_i) dont l'équation générale est la suivante :

$$Y = bX + a$$

$$\text{Avec : } b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2}$$

$$a = \bar{y} - b\bar{x}$$

Avec \bar{x} et \bar{y} les moyennes respectives des x_i et y_i

L'équation de la droite de régression est obtenue par la méthode de moindre carré. De plus, l'outil calcule le coefficient de corrélation et les coordonnées du point moyen $G(\bar{x}, \bar{y})$

2.3.3 SIMULATION NUMÉRIQUE

La démarche à effectuer est la suivante :

- Entrer la série de nombres (x_i, y_i) qui sont séparés par un point-virgule ;
- Entrer la valeur pour laquelle vous souhaitez extrapoler ;
- Afficher le quadrillage.

2.3.4 ETUDE DE CAS POUR LES ÉTABLISSEMENTS D'ENSEIGNEMENT SUPÉRIEUR ET UNIVERSITAIRE**2.3.4.1 TABLEAU RÉCAPITULATIF DES DONNÉES OBSERVÉES**

Premier cas : secteur public

Tableau 1. Les données observées

A. ENSEIGNEMENT ISP				
Périodes	Ets	Etudiants	Personnel	
			Enseignants (PAS)	PATO
2008-2009	79	38232	4796	3928
2010-2011	138	60536	4618	4986
2011-2012	120	53552	4659	5709
B. ENSEIGNEMENT ZST				
2008-2009	150	135927	5033	6915
2010-2011	247	139097	5988	8617
C. UNIVERSITE IFAC				
2008-2009	72	119897	6313	7000
2010-2011	65	163098	5310	5738
2011-2012	43	183863	5470	6588

2.3.4.2 FORMALISME DES EQUATIONS DE BASE

- Pour les établissements : $Y_t = 20,5X_t + 71,33$
- Pour les étudiants : $Y_t = 7660X_t + 35453$
- Pour les enseignants Pas : $Y_t = 431,5X_t + 3494,7$
- Pour le Pato : $Y_t = 890,5X_t + 3093,3$

2.3.4.3 RÉCAPITULATIF PRÉVISIONNEL CI-DESSOUS

Tableau 2. Récapitulatif prévisionnel des données calculées

ANNEES		ETABLISSEMENTS	ETUDIANTS	PAS ENSEIGNANTS	PATO
	RECONSTITUTION	DES	DONNEES		
2008-2009	Y _{t1}	91,83	43113	3926,2	3983,8
2009-2010	Y _{t2}	112,33	50773	4357,7	487,3
2010-2011	Y _{t3}	132,83	58433	4789,2	5764,8
2011-2012	Y _{t4}	153,33	66093	5220,7	6655,3
2012-2013	Y _{t5}	173,83	73753	5652,2	7545,8
2013-2014	Y _{t6}	194,33	81413	6083,7	8436,3
2014-2015	Y _{t7}	214,83	89079	6515,2	9326,3
2015-2016	Y _{t8}	235,33	96733	6946,7	10217,3
2016-2017	Y _{t9}	255,83	104393	7378,2	11107,8
2017-2018	Y _{t10}	276,33	112053	7809,7	11998,3
PREVISION					
2018-2019	Y _{t11}	296,83	119713	12124,7	12888,8
2020-2021	Y _{t12}	317,33	127373	16439,7	13779,3
2022-2023	Y _{t13}	337,83	135033	20754,7	14669,8
2023-2024	Y _{t13}	358,33	142693	25069,7	15560,3
2024-2025	Y _{t14}	378,83	150353	29384,7	16450,8
2025-2026	Y _{t15}	399,33	158013	38014,7	17341,3
2026-2027	Y _{t16}	419,83	165673	42329,7	18231,8
2027-2028	Y _{t17}	440,33	173333	46644,7	19122,3
2028-2029	Y _{t18}	460,83	180993	50959,7	20012,8
2029-2030	Y _{t19}	481,33	188653	55274,7	20903,3
2030-2031	Y _{t20}	501,83	196313	59583,7	21793,8
2031-2032	Y _{t21}	522,33	203973	63904,7	22684,3
2032-2033	Y _{t22}	542,83	211633	68219,7	23574,8
2033-2034	Y _{t23}	563,33	219293	72534,7	24465,3
2034-2035	Y _{t25}	583,83	226953	76849,7	25355,8
2035-2036	Y _{t26}	604,33	234613	81164,7	26246,8
2036-2037	Y _{t27}	624,83	242273	85479,7	27136,8
2037-2038	Y _{t28}	645,33	249933	89799,7	28027,3
2038-2039	Y _{t29}	665,83	257593	94109,7	28917,8
2039-2040	Y _{t30}	686,33	265253	98424,7	29808,3

3 LES GENERALITES SUR LES SYSTEMES DECISIONNELS [5]

3.1 DÉFINITION

Un système décisionnel est un système d'information dédié aux applications décisionnelles tout simplement. IL regroupe un ensemble d'informations et d'outils mis à la disposition des décideurs pour supporter de manière efficace la prise de décision.

3.2 ETUDE COMPARATIVE ENTRE SYSTEME OLTP ET SYSTEME DECISIONNEL OLAP

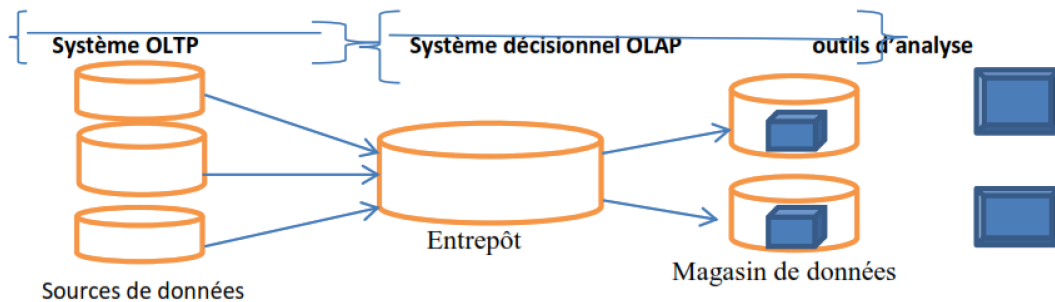


Fig. 2. Architecture des systèmes décisionnels [6]

3.3 DESCRIPTION

- **Les sources de données**, qui sont nombreuses, peuvent être internes notamment les de production ou externes comme Internet, les bases des partenaires à l'entreprise ;
- **L'entrepôt de données** est le lieu de stockage centralisé des informations utiles pour les décideurs. Il met en commun les données provenant des différentes sources et conservent leurs évolutions. Cependant, l'intégration de données consiste à résoudre les problèmes d'hétérogénéité des systèmes de stockage des modèles de données, de sémantique de données.
- **Les magasins de données** sont des extraits de l'entrepôt orientés sujet. Les données sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise ;
- **Les outils d'analyse** permettent de manipuler les données suivant des axes d'analyses. L'information est visualisée au travers d'analyse interactives et fonctionnels dédiés à des décideurs souvent non informaticiens comme directeurs, chefs de services.

3.4 LE TABLEAU 3 : COMPARAISON DES CARACTERISTIQUES DES PROCESSUS OLTP ET OLAP

Tableau 3. Comparaison des processus OLTP et OLAP [7]

	Processus OLTP	Processus OLAP
Données	-Exhaustives	-Résumées
	-Courantes	-Historiques
	-Dynamiques	-Statiques
	-Orientées applications	-Orientées sujets (analyses)
Utilisateurs	-Nombreux	-Peu nombreux
	-Variés(employés, directeurs.....)	-Uniquement les décideurs
	-Concurrents	-Non concurrents
	-Mises à jours et interrogations	-Interrogations
	-Requêtes prédéfinies	-requêtes imprévisibles et complexes
	-Réponses immédiates	-Réponses moins rapides
-accès à peu d'informations	-Accès à de nombreuses informations	

Les problèmes relatifs aux entrepôts de données complexes auxquels nous nous sommes intéressés pour proposer une solution optimale d'usage de l'approche de conception hybride des outils d'aide à la décision basés sur DataWarehouse pour la gestion intégrée des ressources des EESU en RDC plus particulièrement la ville de Kinshasa.

Ainsi, les problèmes abordés constituent effectivement les différents verrous scientifiques du système décisionnel du 21ème auxquels nous avons tenté d'apporter personnellement des solutions optimales adaptées à l'environnement ciblé conformément à l'évolution des bases de données complexes décisionnelles constituant un Méta modèle entrepôt de données objet-classe.

A cet effet, le point focal de nos travaux est la prise en compte du décideur des EESU dans le processus d'entreposage et de partage de données très volumineuses orientées sujets et ou services.

L'originalité de nos travaux de recherche a consisté à concevoir et développer un Méta modèle entrepôt de données pour la gestion intégrée des ressources des EESU dans la ville province de Kinshasa tout en intégrant les techniques de la fouille de données en ligne au sein d'un SGBD Oracle pour traiter des grandes bases et ou entrepôts de données sans limitation de taille de stockage en fournissant des axes d'analyse pertinents dans des intervalles des temps très réduits de l'ordre des quelques secondes.

Ensuite, nous avons montré l'avantage de combiner de la fouille de données avec le Méta modèle entrepôt de données objet-classe en analyse multidimensionnelle pour la création des nouveaux axes d'analyse sémantique plus riches grâce à l'opérateur d'agrégation sémantique du type Roll-Up With K-Means OKITO, permettant de créer un nouveau niveau de granularité dans une hiérarchie de dimension spécifique. K- Means est une méthode de classification automatique qui consiste d'extraire des nouvelles structures sémantiques plus riches pour poser des nouvelles requêtes décisionnelles sur le Méta entrepôt de données regroupant neuf entrepôts des ressources de gestion des EESU. Puis le déployer en toute sécurité en réseau intersites du type backbone à fibre optique avec perspective de connexion à Internet à haut débit

Enfin, en perspective, expérimenter en statistique la droite de régression pour estimer les effectifs des étudiants de chaque année académique pour la construction des infrastructures au-delà de l'horizon 2090 par exemple.

4 ETAT DE L'ART DES APPROCHES DE CONCEPTION DES OUTILS D'AIDE A LA DECISION [8]

D'après nos investigations, nous avons pu identifier les approches de conception ci-après :

- **Entrepôts de données peu centrés décideur** où celui-ci n'est pas forcément en mesure de résoudre, a fortiori des problèmes complexes des décideurs. Ainsi, pour résoudre ce problème, la création des magasins de données représente une première solution possible et tentent de se rapporter des besoins décideurs en fonction de leurs métiers spécifiques ;
- **Modèles d'entrepôts en étoile** : on prend en compte les nouveaux besoins d'analyse des décideurs grâce à l'OLAP mais malheureusement limité en vue d'exploiter des nouveaux axes d'analyse non attendus a priori ;
- **Analyse OLAP limitée aux entrepôts de données**, ces derniers représentent des supports appropriés pour les applications d'analyse en ligne mais une vision multidimensionnelle permet de créer des vues multidimensionnelles de données appelées **cubes de données** dont la vocation est l'analyse en ligne dite OLAP ;
- **Avènement de données complexes** qui sont constitués des textes, image, son, vidéo, base de données, etc., provenant des sources diverses respectivement des données de production, scanners, satellites, enregistrement à distance.
- **Modèles multidimensionnels** sont construits soit sur un modèle relationnel, soit sur un modèle physique réellement multidimensionnel. Selon leur génération respective, il existe trois catégories d'approche des modèles multidimensionnels respectivement ROLAP, MOLAP et HOLAP. Ce dernier demeure le meilleur parmi eux car il représente un système hybride de compensation des insuffisances de deux premiers cités.

4.1 ARBRES DE DECISION

Ils représentent des outils d'aide à la décision sous la forme graphique faisant apparaître les résultats possibles en fonction des choix effectués à chaque étape. Ce sont des algorithmes, méthodes, de segmentation qui ont pris essentiellement leur essor dans le cadre des approches d'apprentissage automatique dite Machine Learning en Intelligence Artificielle. Cette dernière met l'accent sur la convivialité et l'intelligibilité ou la lisibilité des résultats en classification supervisée notamment la sortie des résultats sous la forme des règles logiques de classification à savoir : "Si tel ensemble de conditions sur telles

variables est satisfait **ALORS** le cas appartient à telle classe ». A cet effet, les particularités de l'Intelligence artificielle sont les suivantes :

a) **Les exemples des arbres de décision** sont : respectivement Inductive décision 3 (ZD3), successeur (C4.5), classification avant régression shi-square automatic interaction (CHAID) et Quick unbiased Efficient Startical Trees (QUEST).

b) **Principes de base des arbres de décision**

Deux phases symbolisent les principes de base des arbres de décision :

1. **Phase1** : La construction même des arbres de décision concerne un ensemble d'apprentissage, de processus de division souvent binaire de l'espace donné en sous régions de + en + pures en termes des classes estimés sur base d'un critère spécifique mais dans le cas des données numériques, les approches des séparations parallèles aux axes versus obliques sont possibles comme indiqué dans la figure8 ci-après.

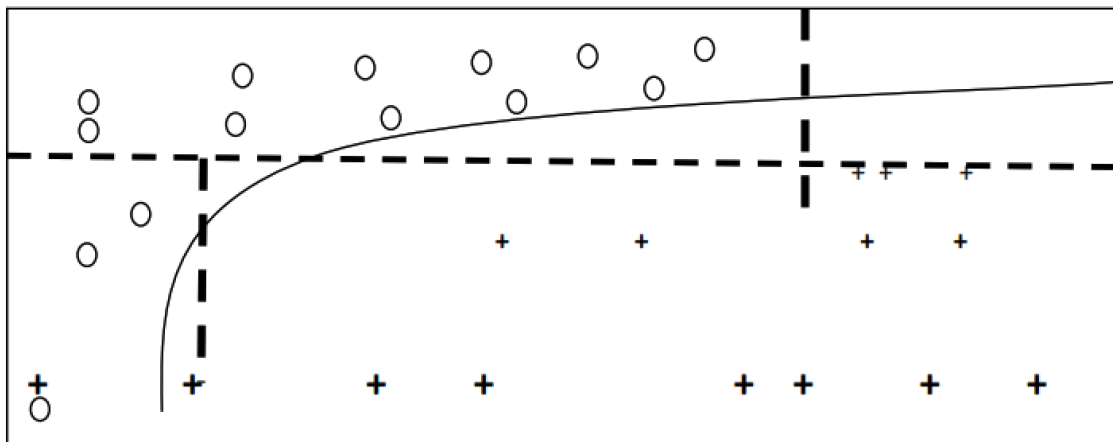


Fig. 3. Séparations parallèles aux axes obliques Root [9]

2. **Phase2** : Elagage ("pruning") consiste à supprimer les branches, parties terminales, représentatives pour garder des bonnes performances prédictives où les nouvelles feuilles labélisées. La décomposition d'un problème de classification en une suite de tests imbriqués porte sur une variable "parallèle aux axes " ou une combinaison linéaire de plusieurs variables oblique.

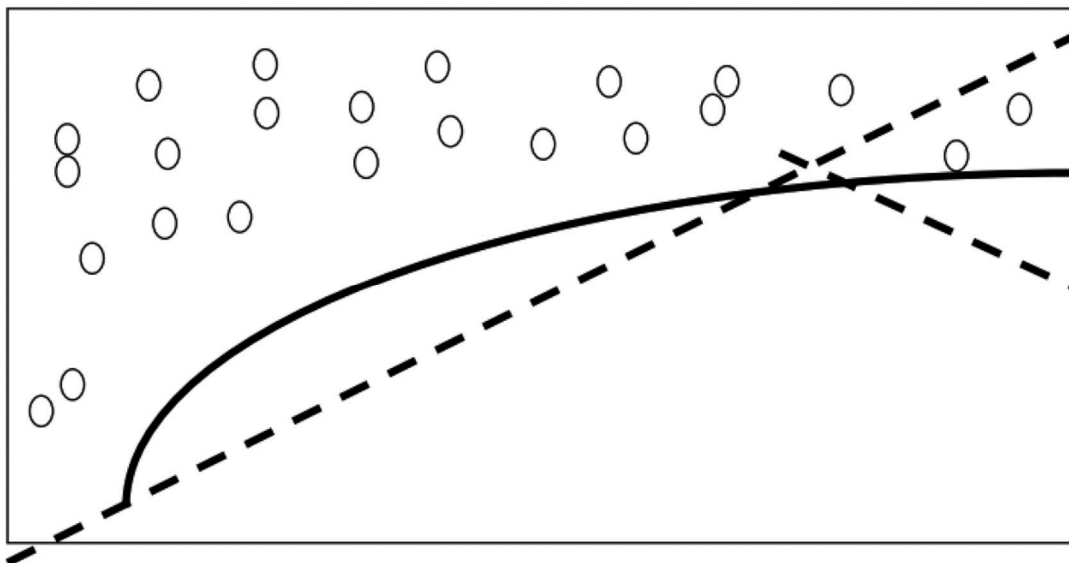


Fig. 4. Schéma des variables parallèles aux axes versus obliques Root [10]

Les règles de classification sous forme d'arbres dont chaque extrémité, appelée feuille, indique l'appartenance à une classe sur les variables mixtes se présente comme suit à titre illustratif avec X= catégorielle tandis que Y et Z sont numériques.

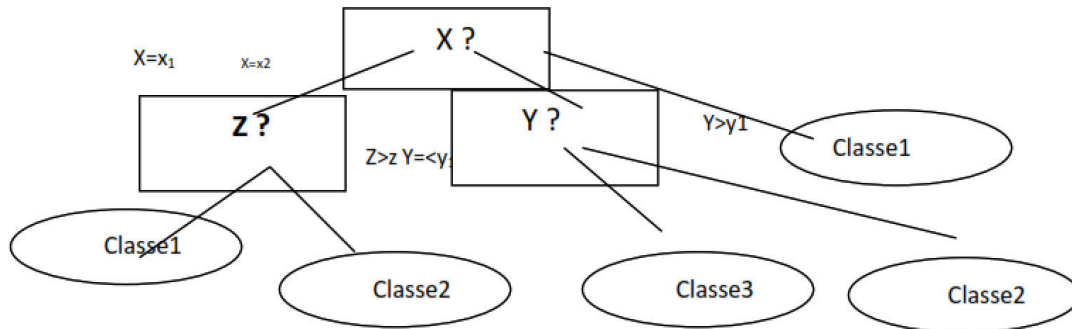


Fig. 5. Schéma illustratif d'un ensemble des exemples positif et négatif

Chaque nœud intermédiaire réalise un test portant sur une variable dont le résultat indique la branche à suivre dans l'arbre. Pour classer un nouveau cas, on est obligé de suivre le chemin à partir de la racine jusqu'à l'endroit souhaité tout en posant des questions successivement.

c) Notion d'Entropie

Le critère de sélection est souvent basé sur la théorie de l'information qui caractérise la notion d'Entropie. Cependant, l'entropie de la distribution des classes est la quantité moyenne d'informations nécessaire pour identifier la classe d'un exemple de E dont voici la formule générale est la suivante :

$$H(E) = -\sum_k p(w_k) \log_2 [p(w_k)] \quad (1)$$

[12]

d) Algorithme des K_voisins les plus proches [13]

C'est une méthode d'apprentissage à base d'instances nécessitant pour des phrases d'apprentissage mais p

5 FOUILLE DE DONNEES EN LIGNE VIA OLAP [14]

Appelée également Data Mining en termes anglo-saxons, elle a été définie en 1996 comme étant un ensemble de techniques d'exploration des données permettant d'extraire d'une base de données des connaissances sous la forme des modèles de description afin de décrire le comportement actuel des données et prédire le comportement futur des données.

5.1 CONTRIBUTION METHODOLOGIQUE

En ce qui concerne l'aspect méthodologique, nos travaux se situent dans la perspective des problèmes d'optimisation des requêtes d'interrogation et la restitution des réponses grâce à la technique de TUNING.

5.2 CONTRIBUTION APPLICATIVE

C'est dans le cadre du projet de la gestion des EESU que se situe notre contribution majeure en termes d'application de techniques de fouille de données en expérimentant l'arbre de décision de la gestion des EESU selon l'organisation hiérarchique des conseils d'administration respectifs.

5.3 ARCHITECTURE TYPE DE DATA MINING

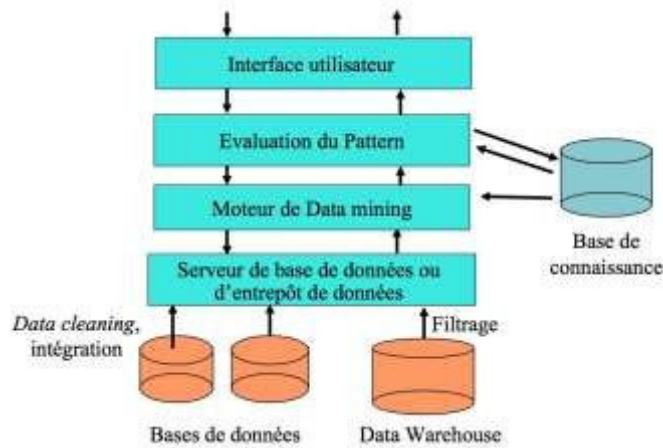


Fig. 6. Architecture de data mining [15]

6 UNE APPROCHE DE CONCEPTION DES OUTILS D'AIDE A LA DECISION BASES SUR DATAWAREHOUSE

Elle représente notre nouvelle approche mixte conçu qui découle des approches existantes étudiées

6.1 SCHÉMA CONCEPTUEL FONCTIONNEL

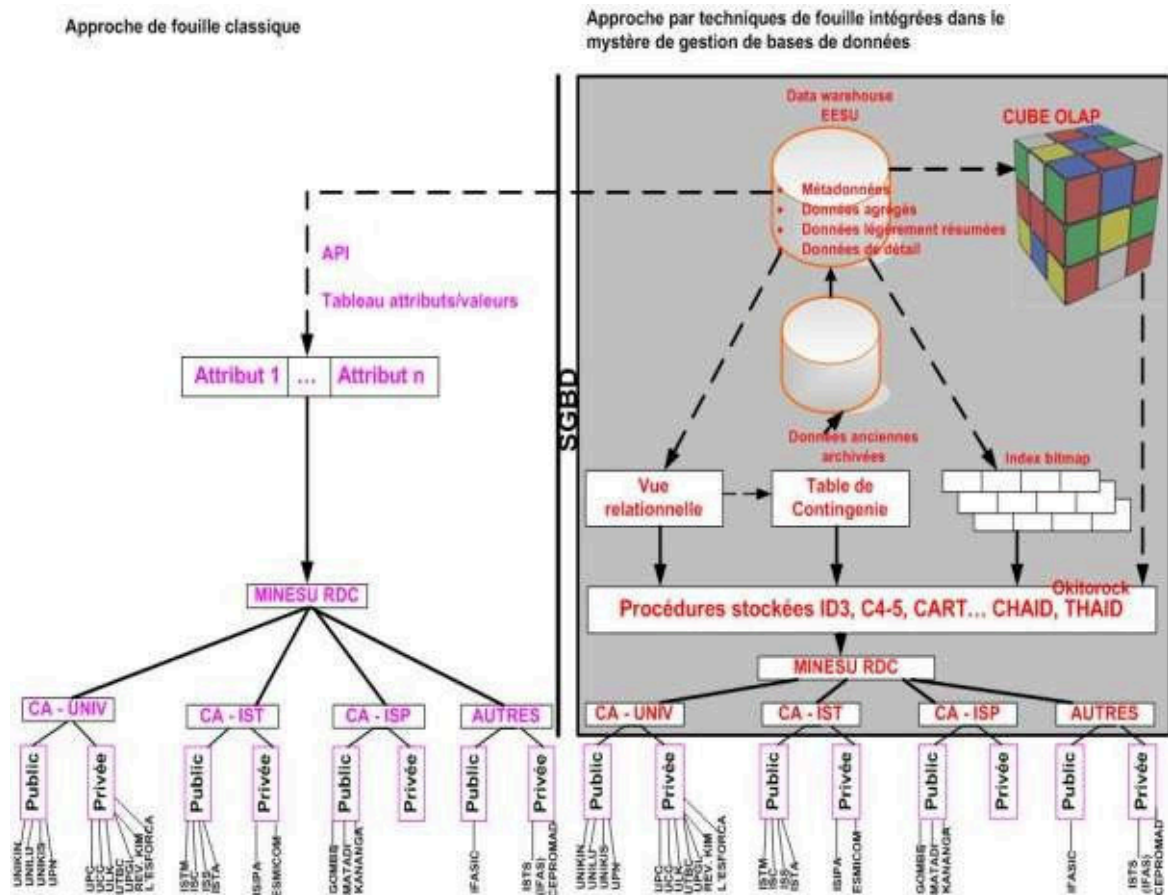


Fig. 7. Approche de conception des outils d'aide à la décision

6.2 DESCRIPTION

6.2.1 IMPLÉMENTATION DU MÉTA MODÈLE ENTREPÔTS OBJETS-CLASSE EESU

Pour valider notre Méta modèle entrepôts objet-classe de données complexes, un prototype a été développé en WINDEV 22 pour représenter un système centralisé de la gestion des ressources des Etablissements d'Enseignement Supérieur et Universitaire dont voici respectivement le schéma fonctionnel et les interfaces conçues ci-dessus.

Point de vue de leur organisation et fonctionnement actuel, tout en présentant la carte de sites universitaires, la répartition du personnel selon les grades et types d'enseignements des étudiants de chaque institution ESU et y la capacité d'accueil des infrastructures.

6.2.2 SCHÉMA FONCTIONNEL DE META ENTREPÔTS DE DONNÉES CONÇUES

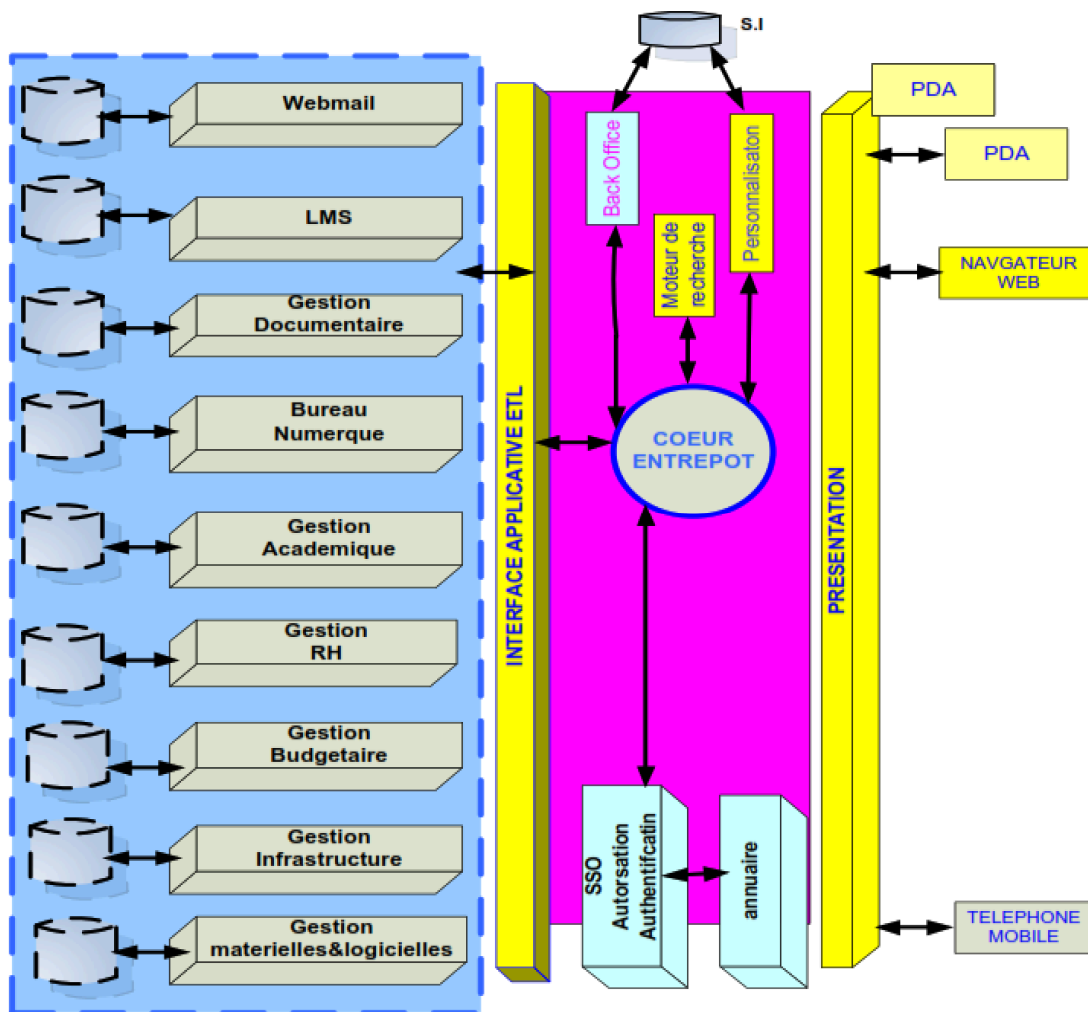


Fig. 8. Méta Entrepôts de données ressources EESU [16]

6.2.3 DESCRIPTION

- **Webmail** est une interface web de messagerie permettant de lire, gérer et envoyer des E-MAILS depuis un browser Internet en mode software avec une connexion SSL sécurisée. Il rend possible l'émission, la consultation et la manipulation des courriels électroniques. Exemple Gmail représente une interface web pour accéder à ses courriels. [17]
- **LMS** qui signifie Learning management system, est un système logiciel orienté web développé pour accompagner toute personne impliquée dans un processus d'apprentissage dans sa gestion de parcours pédagogique. [18]

- **Gestion documentaire:** Elle représente l'ensemble des processus dans un environnement donné permettant de gérer les documents sur support solide ou dématérialisé. [19]
- **Bureau numérique:** il s'agit d'un ensemble intégré de services numériques choisi organisé et mis à disposition de la communauté toute entière. Exemple : Gestion électronique des documents.
- **Gestion académique:** Elle s'occupe de la gestion des matières enseignées, les charges horaires, les frais académiques et l'évaluation des compétences des enseignants selon leur niveau respectif.
- **Gestion de ressources humaines:** Elle est l'ensemble des pratiques mises en œuvre pour administrer le personnel, mobiliser et développer les ressources humaines impliquées dans l'activité d'une organisation spécifique. [20]
- **Gestion budgétaire:** Issu du domaine de la finance, elle désigne l'ensemble des éléments qui se rapportent au budget notamment des entrées et des dépenses d'argent prévues pour une période donnée.
- **Gestion infrastructure :** Elle représente l'ensemble des éléments interconnectés qui fournissent le cadre pour supporter la totalité de la structure respectivement les bâtiments, des laboratoires, les aménagements des routes, des territoires et surtout l'urbanisme des transports nécessaires pour mener bien des activités de développement d'une institution spécifique.
- **Gestion matérielle/logicielle:** Elle concerne la gestion de partitionnement logiciel et matériel en vue de répartir les tâches de système informatique celles qui seront exécutées par un logiciel classique et celles qui seront réalisées par un matériel spécifique. [21]

7 DEVELOPPEMENT DU LOGICIEL A WINDEV 20

Windev 20 a été utilisé comme une plateforme de développement du logiciel de gestion des établissements d'enseignement supérieur et universitaire. Pour ce faire, il était question de concevoir les différentes interfaces intervenantes et leurs codes exécutables correspondants.

7.1 LES PRINCIPALES INTERFACES D'APPLICATIONS



Fig. 9. Interface chargement applications



Fig. 10. Interface Authentification

7.2 LISTES DES CODES



Fig. 11. Interface Menu général

Fen_chargement
 Initialisation de fen_chargement
 Pour i= 1 à 100
 Jauge progresse=i
 P=i Multitaches(5) Fin
 Utilise(fen_connexion) Fenêtre connexion
 Clique sur btn_connexion

Si Sai_user= "okito" et Mot de passe="1234"

Utilise (fen_Menu) Sinon

Info ("Nom d'utilisateur ou Mot de passe incorrect") User=""

Mot de passe="" Donnefocus (user) Fin

7.3 MENU GEST EESU

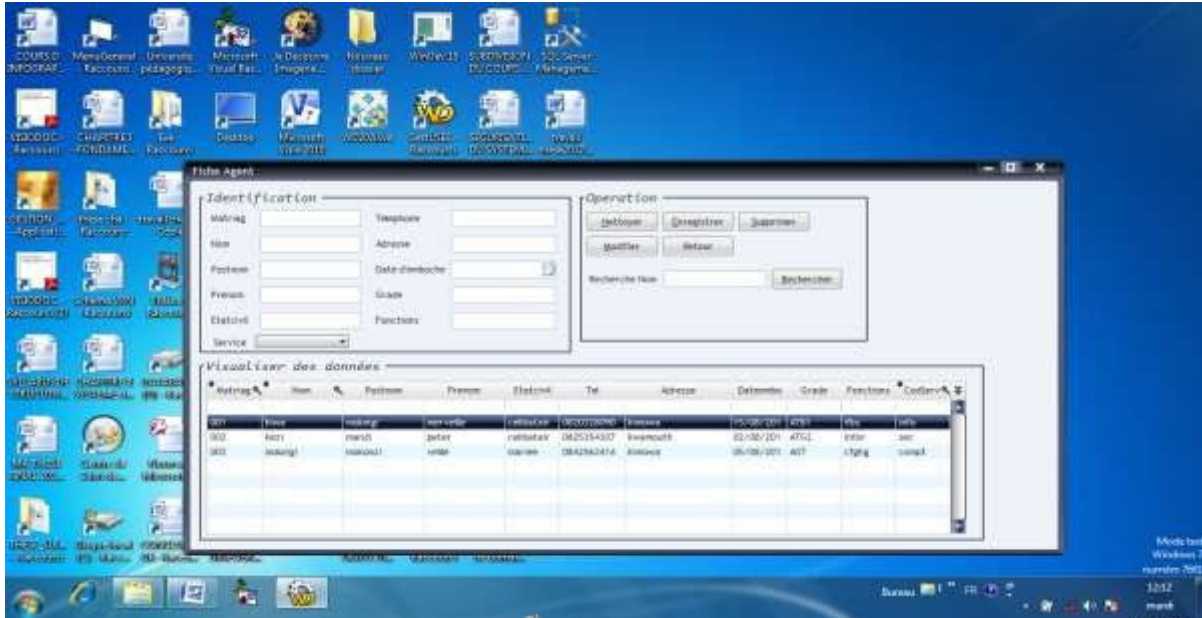


Fig. 12. Interface Gestion desressources humaines

Fen_Agent

Clique sur btnnettoyer

Raz(fen_Agent)

Clique sur btn Enregistrer Ecran vers fichier(Agent) Hajoute(agent) Tableaffiche(tableagent) Raz() Donnefocus(sai_matriag)

Clique sur btnsupprimer Ecran vers fichier(agent) Hsupprime(agent) Tableaffiche(agent)

Clique sur btn Modifier Ecran vers fichier(agent) Hmodifie(agent) Tableaffiche(agent)

Clique sur btn Annuler

Utilise(fen_menu)

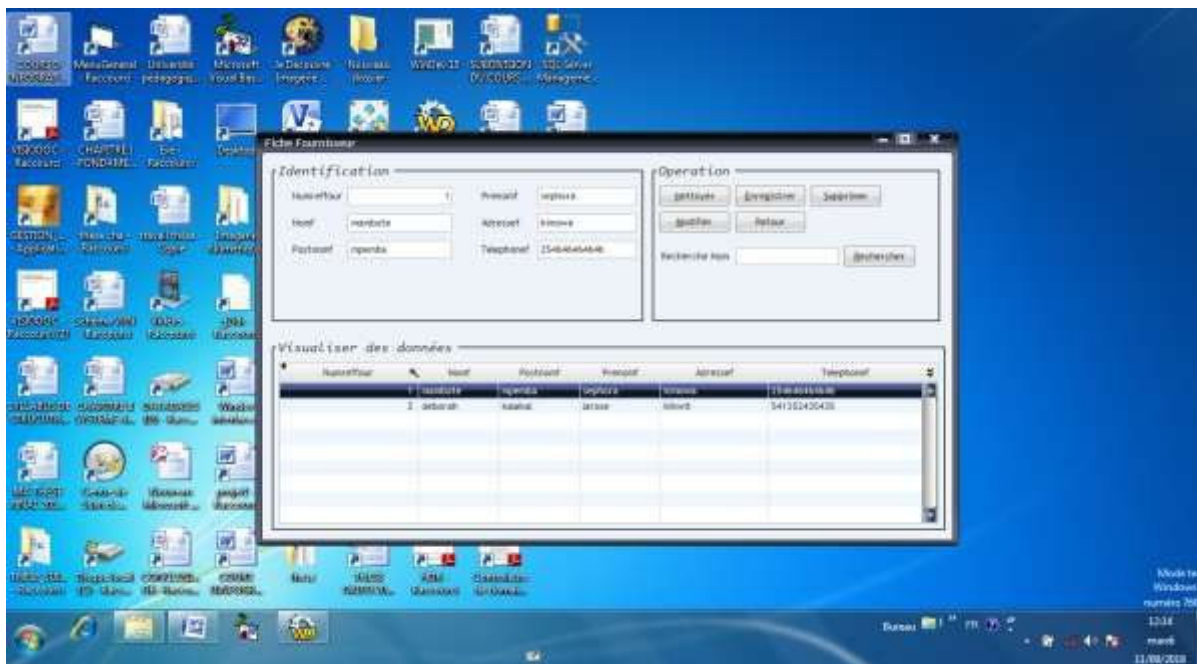


Fig. 13. Interface Gestion de Matériel/Logiciel

Fen_fournisseur

Clique sur btnnettoyer

Raz(fen_nettoyer)

Clique sur btn Enregistrer Ecran vers fichier(fournisseur) Hajoute(fournisseur) Tableaffiche(tablefournisseur) Raz() Donnefocus(sai_numfour) Clique sur btnsupprimer

Ecran vers fichier(fournisseur) Hsupprime(fournisseur) Tableaffiche(fournisseur) Clique sur btn Modifier

Ecran vers fichier(fournisseur) Hmodifie(fournisseur) Tableaffiche(fournisseur) Clique sur btn Annuler Utilise(fen_menu)

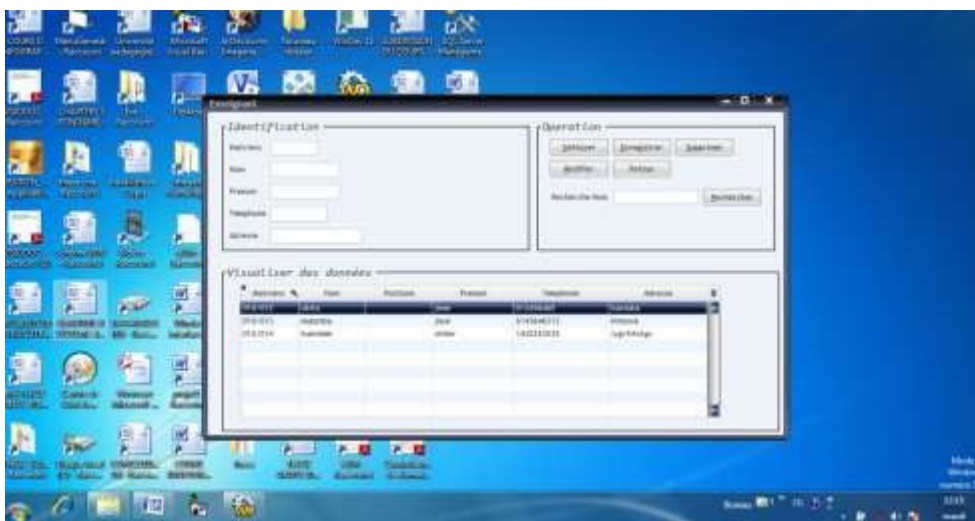


Fig. 14. Interface Gestion Académique

Fen_Enseignant

Clique sur btnnettoyer Raz(fen_enseignant) Clique sur btn Enregistrer

Ecran vers fichier(enseignant) Hajoute(enseignant) Tableaffiche(tableenseignant) Raz() Donnefocus(sai_matriens) Clique sur btnsupprimer

Ecran vers fichier(enseignant) Hsupprime(enseignant) Tableaffiche(enseignant) Clique sur btn Modifier

Ecran vers fichier(enseignant) Hmodifie(enseignant) Tableaffiche(enseignant) Clique sur btn Annuler Utilise(fen_menu)

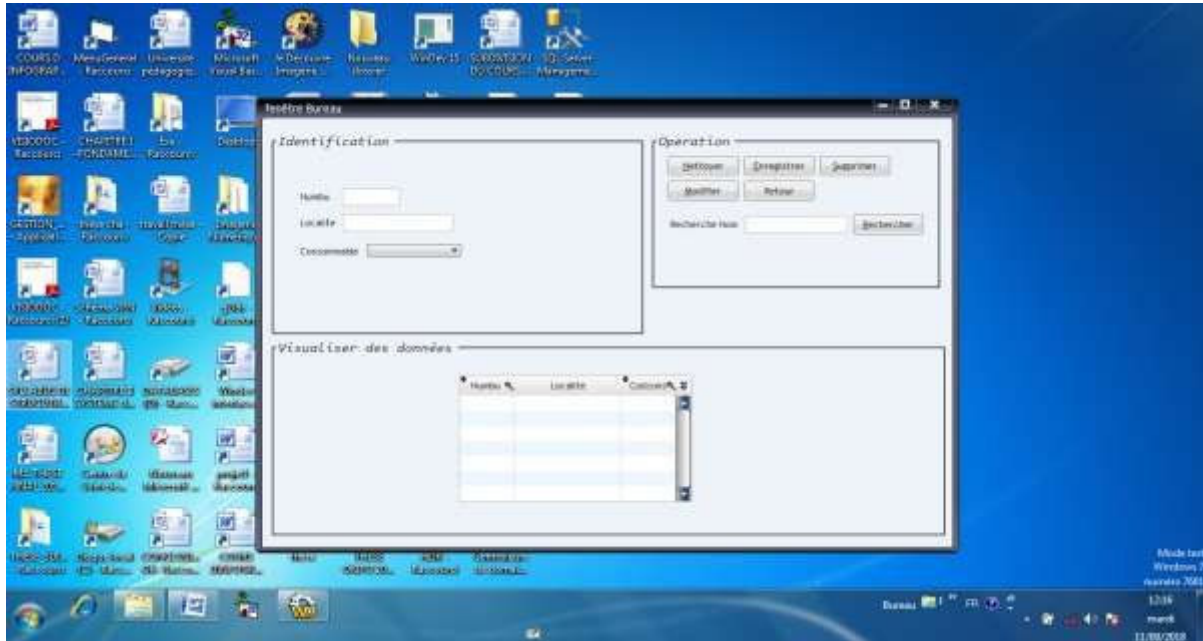


Fig. 15. Interface Bureau Numérique

Fen_bureau

Clique sur btnnettoyer

Raz(fen_bureau)

Clique sur btn Enregistrer Ecran vers fichier(bureau) Hajoute(bureau) Tableaffiche(tablebureau) Raz() Donnefocus(sai_numbu) Clique sur btnsupprimer Ecran vers fichier(bureau) Hsupprime(bureau) Tableaffiche(bureau) Clique sur btn Modifier Ecran vers fichier(bureau) Hmodifie(bureau) Tableaffiche(bureau) Clique sur btn Annuler Utilise(fen_menu)

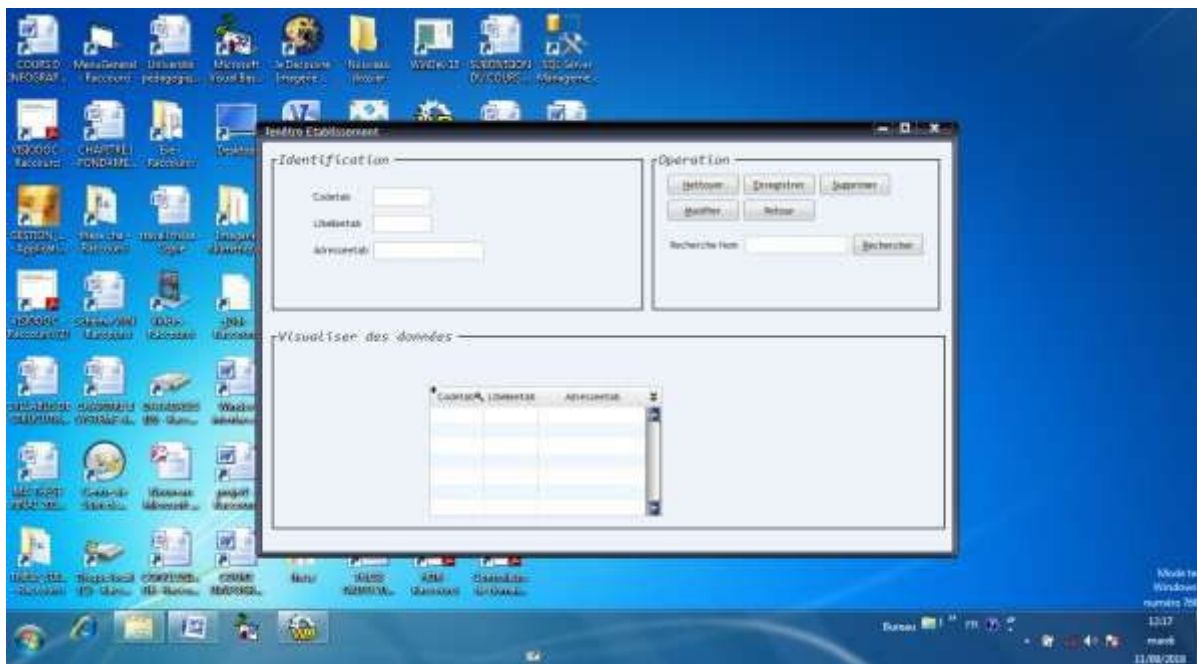


Fig. 16. Interface Gestion budgétaire

Fen_Etablissement Cliquez sur btnnettoyer Raz(fen_etablissement) Cliquez sur btn Enregistrer
 Ecran vers fichier(etablissement) Hajoute(etablissement) Tableaffiche(tableetablissement) Raz()
 Donnefocus(sai_codetab) Cliquez sur btnsupprimer
 Ecran vers fichier(etablissement) Hsupprime(etablissement) Tableaffiche(etablissement) Cliquez sur btn Modifier
 Ecran vers fichier(etablissement) Hmodifie(etablissement) Tableaffiche(etablissement) Cliquez sur btn Annuler
 Utilise(fen_menu)

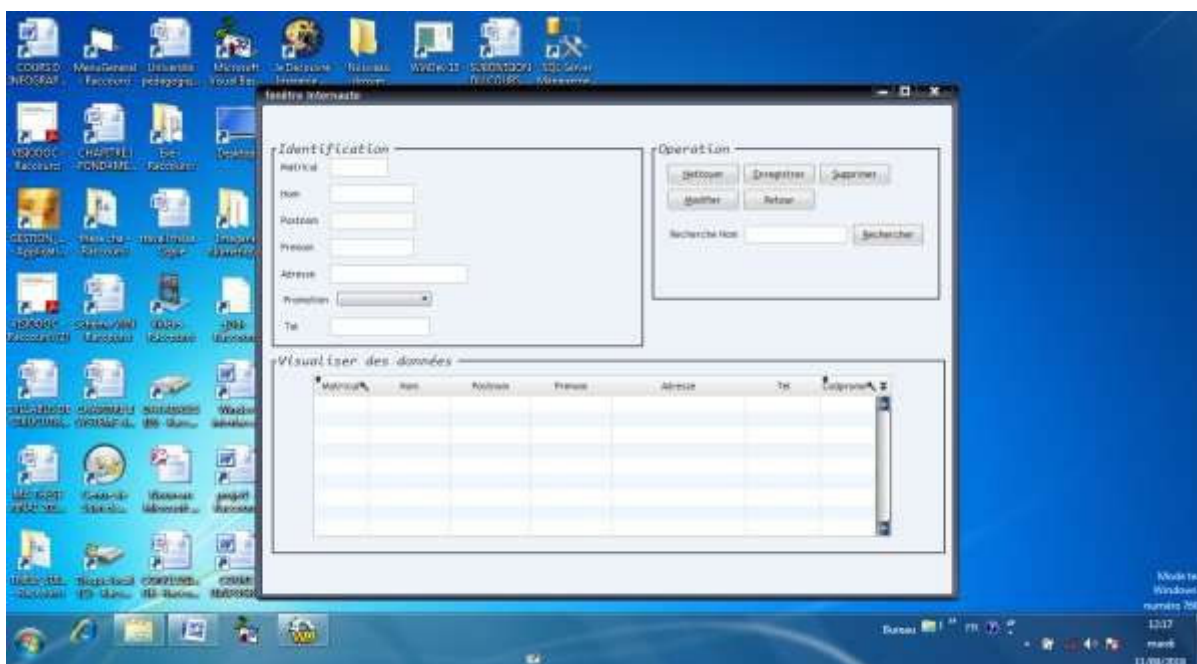


Fig. 17. Interface Gestion LMS

Fen_Internaute

Clique sur btnnettoyer Raz(fen_internaute) Clique sur btn Enregistrer

Ecran vers fichier(internaute) Hajoute(internaute) Tableaffiche(tableinternaute) Raz() Donnefocus(sai_matricul) Clique sur btnsupprimer

Ecran vers fichier(internaute) Hsupprime(internaute) Tableaffiche(internaute) Clique sur btn Modifier

Ecran vers fichier(internaute) Hmodifie(internaute) Tableaffiche(internaute) Clique sur btn Annuler Utilise(fen_menu)

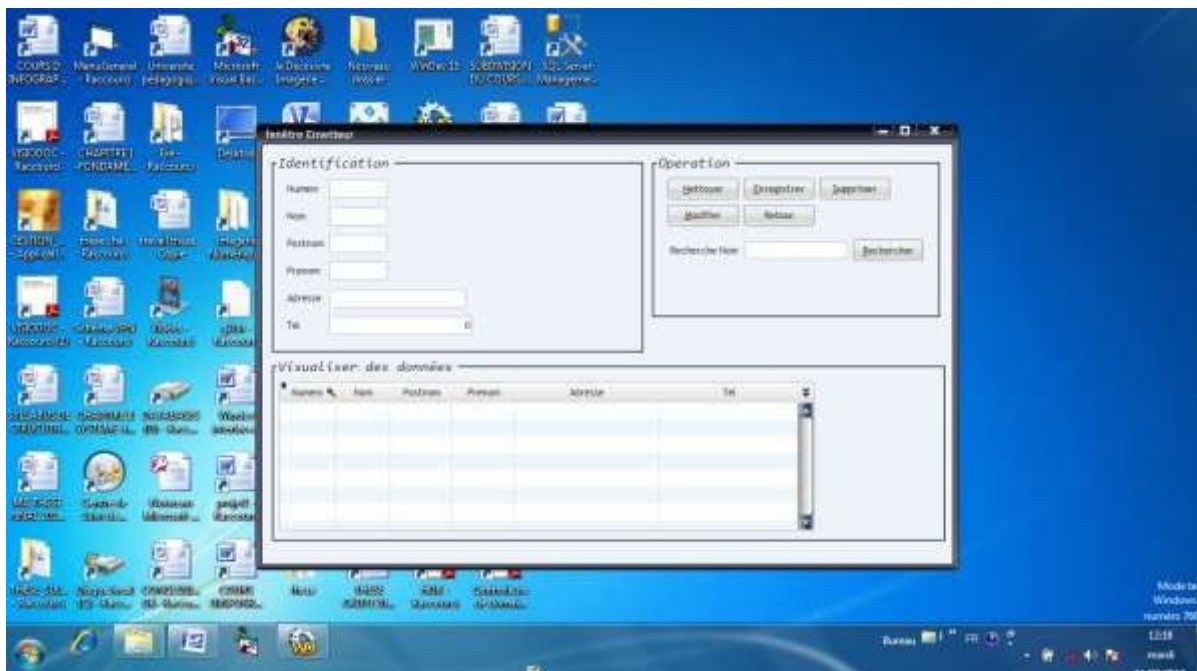


Fig. 18. Interface Gestion Web Mail

Fen_Emetteur

Clique sur btnnettoyer

Raz(fen_emetteur)

Clique sur btn Enregistrer Ecran vers fichier (emetteur) Hajoute (emetteur) Tableaffiche (tableemetteur) Raz() Donnefocus (sai_numem) Clique sur btnsupprimer Ecran vers fichier (emetteur) Hsupprime (emetteur) Tableaffiche (emetteur) Clique sur btn Modifier

Ecran vers fichier(emetteur) Hmodifie(emetteur) Tableaffiche(emetteur) Clique sur btn Annuler Utilise(fen_menu)

8 DEPLOIEMENT DU LOGICIEL DANS UN BACKBONE A FIBRE OPTIQUE EN TOUTE SECURITE

Le présent backbone représente l'environnement où serait partagé le logiciel de gestion des établissements d'enseignement supérieur et universitaire en toute sécurité.

8.1 L'ARCHITECTURE DU BACKBONE

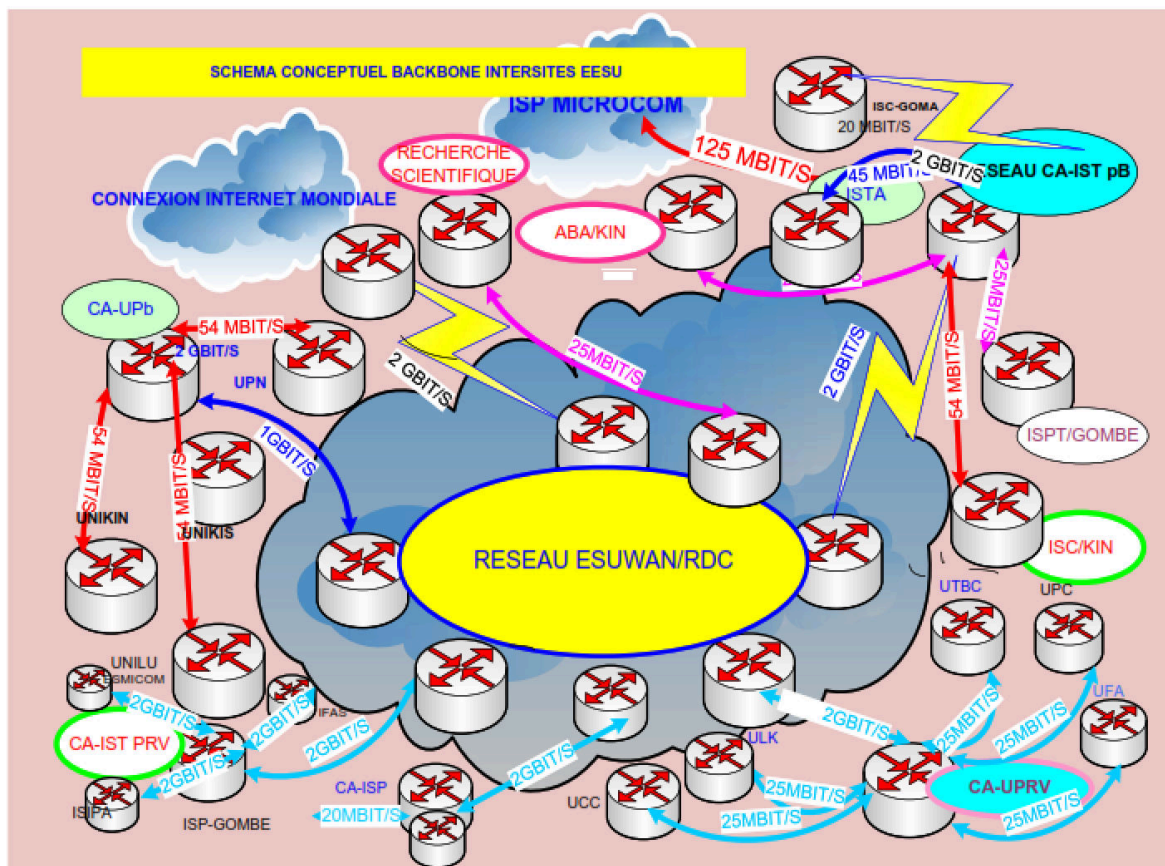


Fig. 19. Backbone intersites EESU réseau wanESU

8.2 INTERNET

C'est ensemble de réseaux interconnectés utilisant tous les mêmes protocoles de montage et de transport TCP/IP en vue de permettre d'accéder à des services notamment E-mail, FTP, les serveurs d'informations en ligne "Web", DNS, etc. Dans son organisation, on distingue les opérateurs, les prestataires de services (ISP), les services et ces outils (protocoles).[22]

8.3 SECURISATION DU META ENTREPOTS EN BACKBONE A FIBRE OPTIQUE (DMZ)

Les solutions de sécurité d'aujourd'hui sont plus hybrides composées de suites d'applications intégrées qui fonctionnent en collaboration pour protéger un réseau privé. Dans ce contexte, il s'agit d'un réseau EESU en backbone à Fibre Optique utilisant des fonctions avancées telles que la journalisation automatique des rapports et les notifications, les pare-feux à des routeurs filtrant qui travaillent ensemble avec des applications proxys pour imposer un contrôle entre l'internet et les backbonde.

8.4 SECURISATION DE BACKBONE EESU A FIBRE OPTIQUE

Pour protéger et contrôler les accès en provenance de l'Internet vers wanESU Backbone intersites EESU à fibre optique, nous avons opté une politique de sécurité par interfaçage architecture hybride des outils de sécurité intégrés notamment VPN, Firewall, serveurs proxy et des logiciels de détection IDS ainsi que IPS comme le montre la figure. [23]

- Virtual private network, VPN en sigle,
- Firewall est un module représentant un ensemble de matériel et logiciel placé en aval de l'internet pour contrôler grâce aux routeurs filtrants des paquets dont il est pourvu, l'accès des utilisateurs non autorisés dans les trafics réseau wanESU

de Backbone intersites EESU. Les filtrants analysent et approuvent les adresses IP autorisés entre Internet et le réseau wanESU.

- Les serveurs proxys se comportent comme les sentinelles qui gèrent tout trafic entre Internet vers l'extérieur malheureusement cela prend beaucoup des temps et coûte cher.
- Les logiciels IDS et IPS s'occupent de détection d'intrusions des virus et des hackers.

8.4.1 INTERFAÇAGE DE SECURITE ENTRE INTERNET ET BACKBONE EESU DMZ

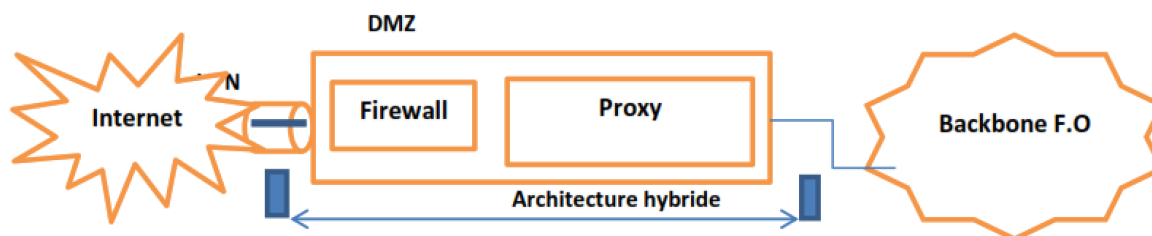


Fig. 20. Interfaçage Architecture hybride entre Internet et Backbone[24] VI.3.3 Cyber sécurité

8.4.2 CYBER SÉCURITÉ

Cyber sécurité est un néologisme désignant le rôle de l'ensemble des lois, politiques, outils, dispositifs, concepts et mécanismes de sécurité, méthodes de gestion des risques, actions, formations et technologies qui peuvent être utilisés pour protéger les personnes et les actifs informatiques notamment matériels et immatériels connectés directement ou indirectement sur le réseau pour assurer la disponibilité, l'intégrité, authenticité, confidentialité, preuve et la non répudiation. [25]

9 CONCLUSION

Les données complexes, des volumes colossaux, renferment beaucoup d'informations multidimensionnelles, qui peuvent être extraites en appliquant des technologies sophistiquées avancées comme la fouille de données et adaptées qu'il faut intégrer dans le processus global d'entreposage piloté par des outils d'aide à la décision. Les nombreux travaux de recherche dans ce domaine montrent la difficulté de trouver un consensus pour entreposer et analyser les données complexes. Nous avons tenté, dans un temps, de décrire un processus complet d'entreposage de données complexes d'aide à la décision tout en soulevant les problèmes liés à leur intégration, leur structuration et leur modélisation multidimensionnelle.

Une nouvelle approche d'intégration de techniques de fouille de données dans le SGBD oracle de modélisation multidimensionnelle et d'analyse en ligne de données a constitué notre conception. Cependant, la technologie Méta modèle objet-classe de données, par usage d'OLAP, est apparue comme une technologie optimale clef pour la gestion centralisée des ressources partagées des Etablissements d'Enseignement Supérieur et universitaire en RDC pour améliorer la qualité des services des décideurs. Cependant, l'analyse OLAP est très limitée puisqu'elle ne permet pas de découvrir des nouvelles structures, d'expliquer un phénomène ou de prédire des nouvelles tendances de construction des infrastructures en fonction de l'évolution croissante des effectifs des étudiants pour chaque année académique. C'est dans ce cadre spécifique que nous nous sommes intéressés à renforcer l'analyse en ligne en la combinant avec les technologies de données intégrée dans un SGBD Oracle pour constituer notre nouvelle approche de conception orientée ressources EESU sans limitation de taille stockage de mémoire en utilisant également la technique de tuning pour optimiser le temps d'accès au Méta modèle entrepôts objet-classe de données.

REMERCIEMENTS

Nous avons l'agréable plaisir de nous acquitter d'un devoir, celui de remercier toutes les personnes qui ont contribué de loin ou de près à la rédaction de cet article. Nos remerciements vont étroitement aux professeurs Engombe Wedi Boniface et Kafunda Katalay Pierre pour leur orientation bénéfique en vue de concrétiser dans la pratique cet article.

REFERENCES

- [1] OKIT'OLEKO, Approche de conception des outils d'aide à la décision basée sur Datawarehouse, UPN, Thèse de doctorat, 2011-2012.
- [2] FABIEN MOUTARDE, Apprentissage Artificiel : Machine-learning, centre robotique (CAOR), Ecole des Mines de Paris, 2011.
- [3] FABIEN MOUTARDE, Apprentissage Artificiel : Machine-learning, centre robotique (CAOR), Ecole des Mines de Paris, 2011.
- [4] BERNARD GRAIS, Droite de régression linéaire, éd. DUNOD, Paris, 1981.
- [5] Franck Ravat, Modèles et outils pour la conception et la manipulation de systèmes d'aide à la décision, Université des Sciences Sociales, Toulouse, France, 2009.
- [6] J.JARKE, M et JEUSFELD., Architecture et Qualité d'Entrepôts de données : Systèmes d'informations, Ed. Dunod Paris 2001.
- [7] IDG., Entrepôts et Analyse complexes centrés utilisateurs : un nouveau défi Lyon II France 2011.
- [8] SERNAT ENCINES ET MARIA TRINIDAD, Entrepôts de données pour l'aide à la décision médicoluée : conception et expérimentation, Université Joseph Fourier, Grenoble, France, 27 Juin 2017.
- [9] Christine Decaestercker, Arbres de décision, Ed. Eyrolles, Paris, 2003.
- [10] Christine Decaestercker, Arbres de décision, Ed. Eyrolles, Paris, 2003.
- [11] Christine Decaestercker, Arbres de décision, Ed. Eyrolles, Paris, 2003.
- [12] Christine Decaestercker, Arbres de décision, Ed. Eyrolles, Paris, 2003.
- [13] YASMINE HANANE ZIGGANE MOKHTAR, Methods de plus proches voisins, éd. Dunod, Paris, 2003.
- [14] MARIUS FIESCHI, data Miningn Université Montpellier, Paris, 2003
- [15] R.LEFEBURE, G.VENTURI., Data Mining : Gestion de la relation client-personnalisation, Ed Dunod Paris 2001.
- [16] R. KIMBALL et Al., concevoir et déployer un entrepot de données, Ed. Eyrolles, Paris, 2000.
- [17] R. KIMBALL, M., Entrepots de données : Guide pratique de modélisation dimensionnelle, ed. Vuibert, Paris, 2003.
- [18] O. BOUISSAID, F. BENTAYEB, « vers l'entreposage de données complexes : structuration, intégration et Analyse. Ingénierie des systèmes d'information », Ed. EDA, Tunisie, 2003.
- [19] J-M FRANCO., « Piloter l'Entreprise grâce au DataWarehouse », Ed Eyrolles Paris 2001.
- [20] WILFRIED DESPAGNE.,« Construction, Analyse et Implémentation d'un modèle de prévision chez opérateur européen du transport et la logique », Université de Bretagne-Sud mai 2012.
- [21] E.BENITZ., « Infrastructure adaptable pour l'évolution des entrepôts de données », PhDthesis, université joseph Fourier, Grenoble, France Septembre 2002.
- [22] STEPHANE LOHIER et D., Internet: services et réseaux, Ed. Dunod, Paris, 2003.
- [23] D. BRENT CHAPMAN et ELIZABETH D. ZWICKY, la sécurité sur l'internet/firewalls, éd. O'Reilly, Paris, 1996.
- [24] DANIELE DROMARD et FETAH, Réseaux informatiques : la sécurité dans les réseaux, cours et exercices, Ed. Dunod, Paris, 2003.
- [25] B. CHESWIK, Firewalls and internet security, repelling the wily Hacker Wesley publishing company, 2004