الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي
Université Abou Bekr Belkaid
Tlemcen Algérie
جامعة أبي بكر بلقايد
تلمسان الجزائر

République Algérienne démocratique et populaire

Université Abou Bakr Belkaid - Tlemcen -

Faculté des sciences
Département informatique

Mémoire de fin d'études

Pour l'obtention du diplôme de master en informatique

Option :système informatique et de connaissances (S.I.C)

THEME

Construction d'un entrepôt de donnée d'identité social

Réalisé par :

Mme : alioui wahiba
Mlle belkhodja souad

Présenté le 12/11/2015 devant le jury composé de MM.

Dr Hnan Abdedjalil

Mm souad khitri

Dr Khellasi abdeldjalil

Mm Souad KhitriMm zeineb Elyebdri ep boukli hasan

Année Universitaire 2017-2018

Sommaire

Introduction

L'utilisation à grande échelle d'Internet, du Web 2.0 et des réseaux sociaux produit instantanément des volumes non habituels de données. Cette explosion de données est une opportunité dans l'émergence de nouvelles applications métiers mais en même temps problématique face aux capacités limitées des machines et des applications.

Notre travail concerne essentiellement de créer un entrepôt de données d'identité sociale pour collecter et stocker toutes les données d'identité sociale, L'objectif de ce projet est d'apporter des solutions aux problèmes posés par les réseaux sociaux dans un environnement décisionnel. Nous nous intéressons en particulier à l'intégration de données dans un entrepôt de données.

Pour créer la base de donnée on commence par l'acquisition des données on utilise des formulaires en PHP et le système de gestion de bases de données MYSQL qui nous permettons d'importer les informations et les données de défèrent source, ensuite on passe à la deuxième étape: Le Stockage où les données sont chargées dans une base de données qui contient quatre tables: table Facebook, table twitter, table Google plus et table temps qui contiennent les identité sociales de chaque réseaux sociale, pouvant traiter des applications décisionnelles. Et enfin la dernière étape : Restitution des données dans cette étapes il existe plusieurs outils, on a choisi l'analyse multidimensionnelle. Dans cette étape nous avons créé un cube OLAP, l'utilisant de l'outil PALO développé par JedoxAG avec une suite de logiciels de BI open source.

Le plan de travail de notre mémoire est comme suite :

1ere chapitre Réseaux sociaux : on a étudié les réseaux sociaux puis on a présenté les réseaux connus (Facebook, Twitter, Google plus).

2eme chapitre Entrepôt de données on a présenté des notions sur les entrepôts de donnée (définition, caractéristiques, Synthèse sur les outils pour les entrepôts de données ….)

3eme chapitre Application : on a présenté la démarche utiliser pour la construction de l'entrepôt de données.

Chapitre I
Réseaux sociaux

I. Introduction

L'internet est un ensemble des ordinateurs avec des files réseaux .ou des nœuds relies par des arcs. Pour que l'information se passe-t-il dans un temps court avec simple méthodes ; La communication entre les personnes dans des divers domaines (social -politique-commerce ….etc.) ; les différents choses qui sont changée entre eux sont : des textes ; des photos ….) ; L'internet rendre le monde un petit village et leur utilisation augment jour par jour. . L'évolution de l'internet arrive avec différents les réseaux sociaux

II. L'arrivée du web 2.0

Le web 2.0 arrive après le web .il vient avec de nouveauté qui était la sémantique des mots et que l'information toquée selon leur sens .alors cela exige des explorations des réseaux sociaux l'un a l'autre dans chaque année qui sont différents (Wikipédia -face book ...) qui sont utilisée par différents catégories des personnes (jeunes …) [1]

III. Définition d'un réseau

Un réseau est un ensemble de nœud ou (pole) relie par des liens ou (canaux) atteindre pour afin d'échanger des informations, de partager des ressources, de transporter de la matière ou de l'énergie dans différents domaines (informatique, télécommunication, énergie…).Les nœuds peuvent avoir des fonctions plus au moins complexes de distribution, de concentration, d'enrichissement, tandis que les canaux assurent une fonction de transport.

Exemples : Réseau routier, réseau électrique, réseau fluvial, réseau de téléphonie, réseau de distribution, réseau d'agences bancaires.

Ces réseaux ont plusieurs types notre étude deviendra sur les réseaux sociaux :

III.1. Définition Réseau social

Un réseau social est un ensemble de personnes, d'associations, d'établissements, d'organismes ou d'entités sociales qui ont le même objectif et qui sont en relation pour agir ensemble dans un espace virtuel par les courriers électronique ou les messages instantané.**[2]**

III.2. Les réseaux sociaux et les sites internet

Les réseaux sociaux et sites Internet ne sont pas opposés. Un client se rend sur un site Internet pour s'informer sur un produit ou un service. Les journalistes sont à l'affût de sources d'information. Les fournisseurs guettent les évolutions et les nouveautés. En bref : le visiteur cherche et trouve dans un site Internet des données, des chiffres, des faits. Sur Facebook ou Google Plus, ils veulent au contraire communiquer, commenter, s'amuser, nouer des relations, suivre des recommandations, etc.

Informations sur un site Internet propre et relationnel sur une page fan sont tous deux nécessaires à la présence harmonieuse sur Internet d'une idée commerciale. Le buzz des réseaux sociaux augmente la pertinence des informations du site Internet, le site Internet restant la carte de visite d'une marque sur la Toile. C'est pourquoi la question ne devrait jamais se poser de choisir entre les deux mais plutôt de les concilier efficacement, pour offrir au client une image de marque aussi uniforme que possible. **[3]**

III.3. Théorie des réseaux

A partit de la théorie des graphe on à la théorie réseaux sociaux qui est un champ extrêmement actif dans le milieu universitaire et plusieurs outils de recherche d'analyse des réseaux sociaux sont disponibles en ligne et sont relativement faciles à employer pour présenter simplement un graphe de réseau social.

IV. Différents types de réseaux:

On a résumée dans le tableau au-dessus.

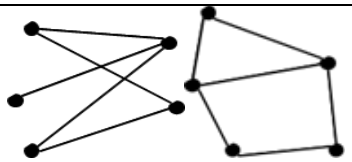
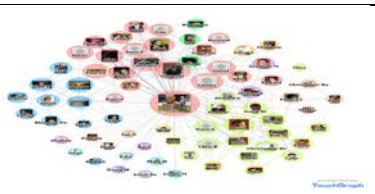
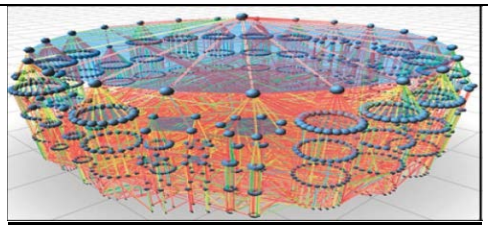
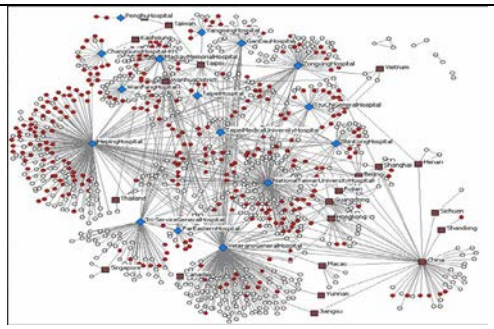
Type réseaux	Exemple	Schéma
Unipartis/ Multipartis :	Réseau d'amitié / Réseau d'achat client-produit	 figure I-1 figure I-2
• Orientés/Non-Orientés:	Réseau de collaboration / Réseau d'appels téléphoniques	
• Avec contenu:	Réseau d'intérêts communs	 figure I-3
• Avec structure relationnelle complexe:	Réseau de relations professionnelles	 figure I-4
• Avec dynamique importante	Réseau de contacts de proximité géographique	 figure I-5

Tableau 1-1 les types des réseaux

V. Structure des réseaux sociaux

V.1. grands types de structure

- **Réseau régulier** : un nombre de liens des nœuds est identique, densité faible, coefficient de clustering élevé, la distribution par un pic pour le degré de marque.
- **Réseau aléatoire :** le lien est résultat d'un processus aléatoire, moyenne faible, distribution des degrés suit une loi de poisson. Coefficient de clustering élevé
- **Réseau petit monde :** Notion populaire: 7 degrés de séparation, Distance moyenne très courte,
- **Réseau scale-free :** Découvert par Barabasi en 1999, il caractérisé par distribution des degrés suit une loi de puissance. [4]

VI. Les principales fonctions des réseaux sociaux

- la création d'une page contenant ces informations personnelles
- la messagerie interne entre les membres
- les groupes de discussion
- la possibilité de mettre en ligne des photos, musiques ou vidéos, des jeux, des quiz
- la possibilité également de créer des listes de contacts
- la possibilité de connaître les dernières actions de ses amis sur le réseau. [5]

VII. Des réseaux sociaux connus

VII.1. face book

Facebook est la plateforme Web favorite des internautes pour les échanges sociaux. Elle a pour but de regrouper les individus entre eux autour de liens et d'intérêts communs, et permet d'échanger du contenu (texte, lien, photo, vidéo). Facebook présente aussi de nombreux services pour les entreprises qui désirent interagir avec leur public, augmentant ainsi le niveau et la fréquence des échanges entre les individus et les marques. La plateforme compte actuellement plus de 500 millions d'utilisateurs à son actif (statistique septembre 2010), dont plus de 50 % qui se connectent plusieurs fois par jour. Les possibilités d'utilisation, la portée de diffusion et les bassins d'intérêts potentiels de Facebook font de cette plateforme un réseau social de choix qui vous

permettra de rejoindre un public présent et actif.Facebook comme outil de Knowledge Management(KM) :

Management des connaissances ou management par les connaissances se sont les méthodes et outils logiciels permettant d'identifier, de capitaliser les connaissances de l'entreprise afin notamment de les organiser et de les diffuser. [6]

La gestion des connaissances est constituée 5 phases, souvent connues sous le terme de « cercle vertueux de la gestion des connaissances »:

- Le repérage des connaissances ;
- La préservation des connaissances ;
- La valorisation des connaissances ;
- La création et le partage des connaissances ;
- L'actualisation des connaissances

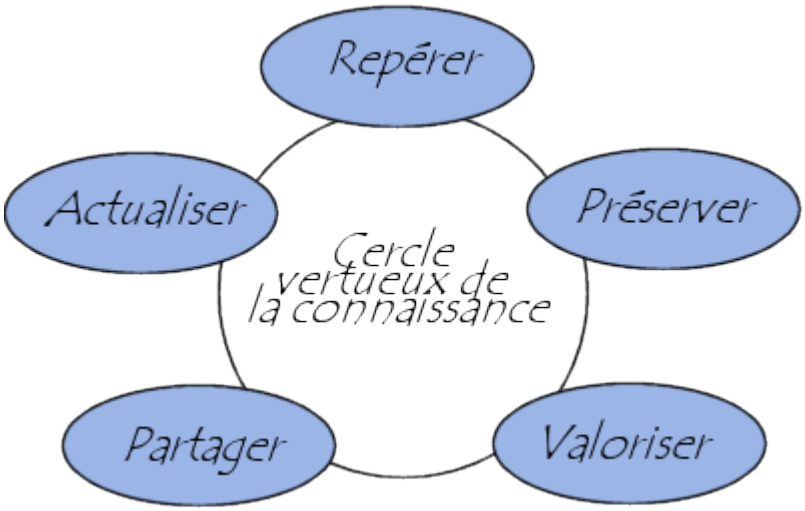


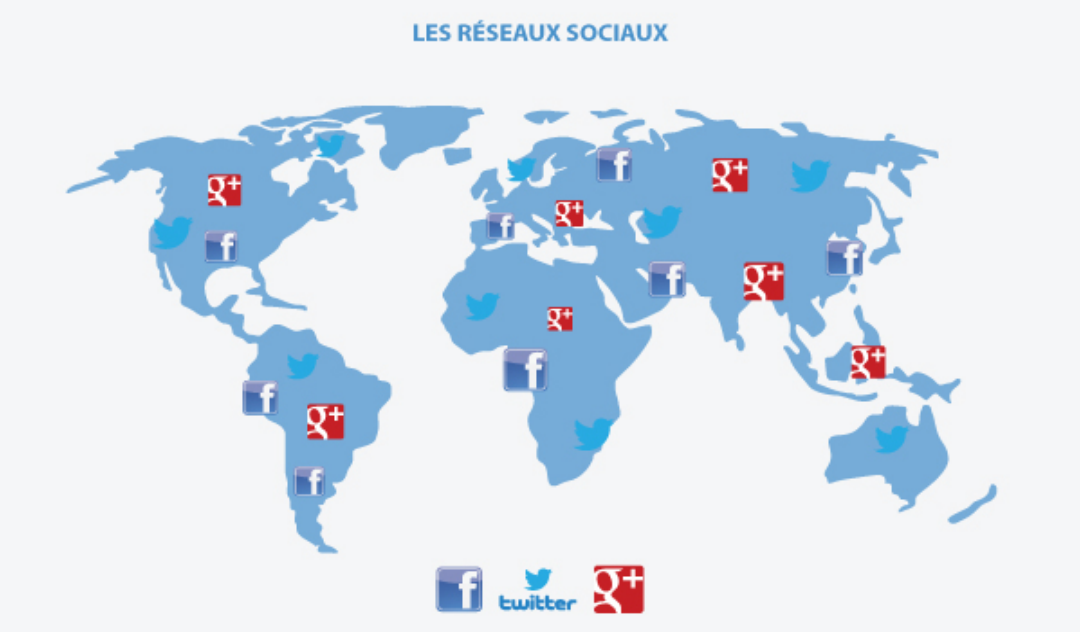
Figure 1-6 cercles vertueux de la gestion des connaissances

VII.2. Twitter

Twitter est une plateforme de " microblogging " qui permet à ses utilisateurs de bloguer grâce à des messages courts de 140 caractères. C'est un réseau social accessible à tous et qui permet de rejoindre rapidement son public par le biais de publications de nouvelles et de liens. Les utilisateurs de Twitter peuvent suivre les activités des entreprises, des marques comme des individus selon leurs goûts et leurs intérêts grâce à une interface simple et facile d'utilisation. En août 2010, on comptait environ 190 millions d'utilisateurs sur Twitter

VII.3.Google+

Est un concurrent de face book, dans une période court de sa devenue Google a réussi à fédérer des millions d'utilisateurs.



FigureI-7 image des réseaux sociaux

VIII. Réseaux sociaux, analyse et data mining

La difficulté pour la fouille de donnée était le modèle a choisir pour le stockage de nombreuse donnée, avec le « Web2.0», la situation change et les exposée de journée permettront au et aux chercheur et praticiens (tacticiens, analistes, exploitant des sites réseaux sociaux) de expliquer et étudier les problèmes du domaine et identifier les avances prochainement et les solutions trouvés.[7]

IX. Applications des réseaux sociaux

Il existe plusieurs types d'applications pour les différents réseaux sociaux :

- Les applications dites ludiques de type jeux concours, quizz
- Les applications collaboratives de partage de contenus

IX.1. Création des applications pour les réseaux sociaux

Pour la première fois, en Août 2011, Facebook a dominé Google en nombre de visiteurs uniques avec plus de 245 millions de visiteurs en un mois. Plus qu'un réseau social, c'est aujourd'hui le point d'entrée exclusif à Internet pour plusieurs centaines de millions d'utilisateurs.

IX.2. L'open Graph 2.0 intégré dans les applications

L'open Graph 2.0, lancé par Facebook, est un protocole qui vous permet d'interagir avec le profil de fan via leurs actions sur d'autres supports Internet. Cette fan écoutée tel morceau, lit tel article, visualise telle vidéo, participe à tel jeu concours…

Les réseaux sociaux n'a des limite ils se trouvent dans tout le domaine :

On d'autre exemple dans la sociologie : la théorie de réseaux sociaux a était étudier sur des échantillons de relation social pour expliquer de nombreux ou divers phénomène de la vie courante science sociale.

Les système sociotechnique sont lies a l'analyses de réseaux et sur les relations parmi les individus, les institutions, les objets et les technologies [8]

X. But des réseaux sociaux dans Les entreprises

- créer une communauté en premier lieu,
- identifier les fans de la marque et en faire des ambassadeurs,
- développer leur notoriété, créer des concours, de l'animation publicitaire,
- mieux s'engager et échanger avec les différents publics de l'entreprise,
- donner un visage humain à l'entreprise,
- recueillir des commentaires sur les produits / services,
- recruter et fidéliser le personnel,
- établir des contacts avec de nouveaux clients,
- organiser des événements,
- mener des études de marché pour mieux connaître les concurrents et recueillir des informations sur les nouveaux produits et nouvelles technologies.

XI. Les perspectives des réseaux sociaux géo localisés

L'utilisateur de Smartphone accèdent aux réseaux sociaux sur cette mobile, la version a doublé, elle était placé dans le premier rang mondial, d'autre réseaux sociaux fait leur apparition et la particularité est intégrer des fonctions de géolocalisation qui permettent au .usager de savoir où se trouvent leur amis et quelle sont les trois lieux préférées[9]

Conclusion

Les réseaux sociaux se sont développés fortement ces dernières années. Nous avons assisté à une forte expansion de leur nombre mais aussi de leur type. Maintenant, chaque internaute peut, en théorie, trouver un réseau social qui lui correspond, qu'il soit à caractère général, thématique ou professionnel. Le développement rapide de ce phénomène a amené les entreprises à participer. Elles sont créées des sites sociaux pour plusieurs activités comme l'animation des annonces …etc.

Alors, les réseaux sociaux est devenue obligatoire pour qu'un utilisateur entré à l'internet.

Les réseaux sociaux sont encore peu utilisés par les compagnies, même si de manière générale, celles-ci prévoient de plus en plus de les utiliser dans leurs stratégies futures.

Chapitre 2
Entrepôt de données

I. Introduction

Les entrepôts des données intègrent les informations en provenance de différentes sources, souvent reparties et hétérogènes ayant pour objectif de fournir une vue globale de l'information aux analystes et aux décideurs.

La construction et la mise en œuvre d'un entrepôt de données représentent une tâche complexe qui se compose de plusieurs étapes. La première à l'analyse des sources de données et à l'identification des besoins des utilisateurs, la deuxième correspond à l'organisation des données à l'intérieur de l'entrepôt. Finalement, la troisième sert à établir divers outils d'interrogation, analyse, de fouille de données.

II. Concepts principaux des entrepôts de données

II.1. Définition :

Entrepôt de données, ce qui se traduit généralement en anglais par Data Warehouse. Ce concept a été formalise en 1992 par Bill Inmon dans l'ouvrage "Developing the Data Warehouse" (Construire l'Entrepôt de Données) de la façon suivante : ≪*Un Data Warehouse est une collection de données thématiques, intégrées, non volatiles et historiées pour la prise de décisions.* ≫ [10]

Un entrepôt de données (Data Warehouse) est une base de données architecturée pas seulement pour un traitement transactionnel des données mais pour des requêtes et des analyses. Ils (entrepôts des données) séparent les opérations analytiques des opérations transactionnelles.

II.2. Caractéristiques

Donc dans un entrepôt selon Bill Inmon, les données sont

- orientées **par sujets** :

Les données organisées par sujet (clients, vendeurs, production, etc.) contiennent seulement l'information utile à la prise de décision.

Les systèmes opérationnels sont plutôt orientés autour des traitements et des fonctions.

• **intégrées** :

Les données, provenant de **différentes sources** (systèmes légués) sont souvent structurées et codées de façons différentes. L'intégration permet d'avoir une représentation **uniforme, cohérente et transparente**. Lorsque les données sont agrégées, il faut s'assurer que l'intégration est correcte.

• **historiques** :

Un entrepôt contient des données "anciennes", datant de plusieurs années, utilisées pour des comparaisons, des prévisions, etc.

• **non volatiles** :

Une fois chargées dans l'entrepôt, les données ne sont plus modifiables. Elles sont uniquement accessibles en lecture (pour l'instant...).

II.3. Différences entre Entrepôts et SGBD

Le tableau 1.1 résume ces différences entre les systèmes de gestion de bases de données et les entrepôts de données

	SGBD	Entrepôts de données
Objectifs	*Gestion et production*	Consultation et analyse
Utilisateurs	*Gestionnaires de production*	Décideurs, analystes
Taille de la base	*Plusieurs gigaoctets*	Plusieurs teraoctets
Organisation des données	*Par traitement*	Par métier
Type de données	*Données de gestion (courantes)*	Données d'analyse (résumées, historisées)
Requêtes	*Simples, prédéterminées, données détaillées*	Complexes, spécifiques, agrégations et *group by*
Transactions	*Courtes et nombreuses, temps réel*	Longues, peu nombreuses

Tab.2.1 – Différences entre SGBD et entrepôts de données

II.4. Construction d'un entrepôt de données

Pour construire un entrepôt de donnée il existe trois phases principales : Acquisition, Stockage et Restitution des données

II.4.1. Acquisition:

- Extraction de données
  - Snapshots ou différentiels des sources
  - Transferts, cryptage, compressions, etc.
  - Objectif : minimum de changement par rapport aux sources
- Transformation
  - Résolution des conflits au niveau du schéma (différents attributs pour la même information)
  - Identification des valeurs et réconciliation
- Nettoyage
  - Élimination des dupliqués, contraintes d'intégrité, etc.
  - Données incomplètes ou absentes
- Chargement
  - Tris, résumés, calculs divers, etc.
  - Pbs: très grand volume de données, efficacité, quand calculer les index et les tables de résumés, reprise sur panne

II.4.2. Stockage :

Les données sont chargées dans une base de données pouvant traiter des applications décisionnelles.

II.4.3. Restitution des données :

Il existe plusieurs outils de restitution (tableaux de bord, requêteurs SQL, analyse multidimensionnelle, data mining ...)

II.5. Architecture générale

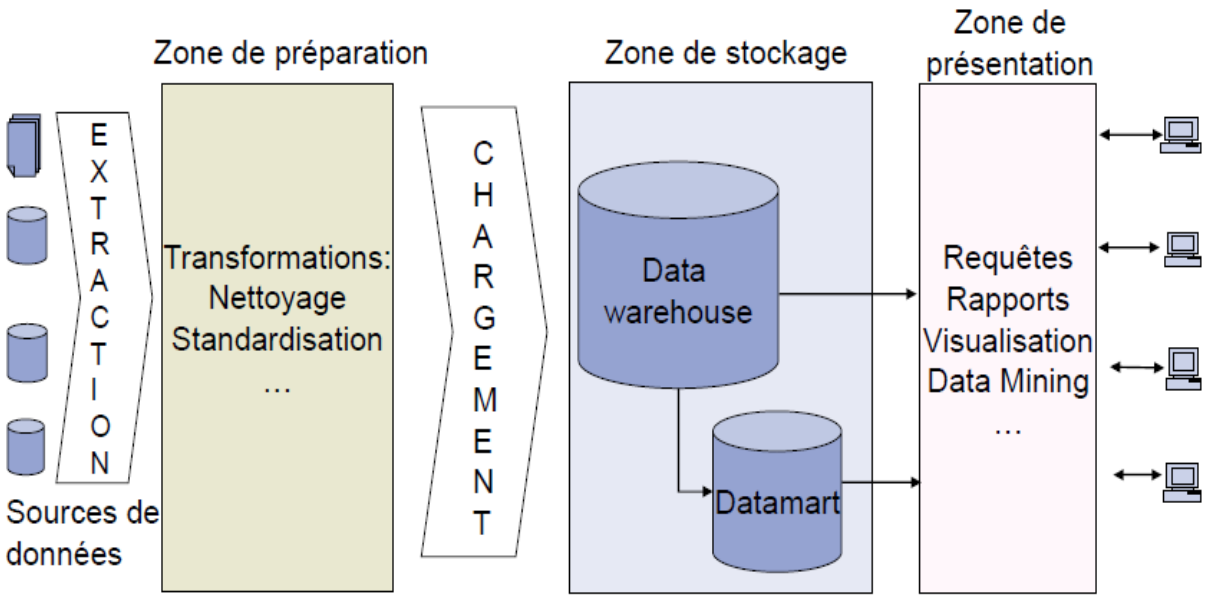


Figure 2.2 Data warhouse

Le premier niveau est un SGBD. Les données sont extraites à partir des bases de données transactionnelles, nettoyées et transformées avec des outils ETL (Extract-Transform-Load ou en français extraction, transformation et alimentation), et intégrées dans l'entrepôt de données. Le SGBD contient aussi un ensemble de métadonnées concernant les sources de données, les mécanismes d'accès, les procédures de nettoyage et d'alimentation, les utilisateurs, etc.

III. Le data Mart

III.1. Définition

Selon Séraphin LOHAMBA OMATOKO un Data Mart est un entrepôt qui stock des données provenant de systèmes opérationnels ou d'autre sources, conçu pour répondre aux besoins spécifiques d'un département ou d'un groupe d'utilisateurs en termes d'analyse, de contenu, de présentation et de facilité d'emploi. Les informations y sont stockées dans un format qui est familier aux utilisateurs. Un Data Mart ressemble en fait à un Data Warehouse sauf qu'il est moins générique. Une approche courante consiste à maintenir des informations détaillées au niveau du Data warehouse et à les synthétiser dans un Data mart pour chaque groupe ou département fonctionnel. Un autre choix de conception consiste à créer des Data marts pour chaque département puis à fusionner

ultérieurement ces données dans l'entrepôt global. Chacune de ces méthodes présente l'avantage de centraliser les informations pour les utilisateurs finaux.

Les caractéristiques propres aux Data Mart sont :

- Les données sont spécialisées pour un groupe ou département particulier ;
- Ils sont conçus pour un accès facile ;
- Le temps de réponse est optimal pour un volume de requêtes moindre ;
- Les différents Data Marts indépendants peuvent être dynamiquement couplé pour se métamorphoser en Data Warehouse ;
- Les Data Marts sont plus flexibles que les Data Warehouse.

En raison de la nature simplifiée et spécialisée des Data Marts, les entreprises choisissent ces magasins de données comme solution rapide à leurs besoins en matière d'aide à la décision. [11]

III.2. Data Warehouse versus Data Mart

Data Warehouse	**Data Mart**
Utilisation globale de l'entreprise	Utilisé par un département ou une unité fonctionnelle
Difficile et plus long à implémenter	Plus facile et rapide à implémenter
Volume de données plus important	Volume de données plus petit et spécialisé
Développé sur la base de données actuelle	Développé sur les bases des besoins utilisateurs

Tab.2.2 – Différences entre Warehouse et Data Mart

Les Data Marts représentent de toute évidence une réponse rapide aux besoins des différents départements de l'entreprise. Leur coût moindre et leur facilité d'emploi permettent une implémentation rapide et un retour à l'investissement presque immédiat. Il faut tout fois être prudent lorsque des Data marts sont ainsi crées pour plusieurs divisions. Ces dernières utilisent souvent des représentations différentes de certains concepts de gestion. Par exemple, les départements finances et marketing peuvent tous deux effectué un suivi des ventes réalisées par l'entreprise, mais défini différemment ce concept. Plus tard, si un employé du marketing a besoin de recueillir certaines

informations à partir du Data Marts des finances, l'entreprise sera confrontée à un problème. Par conséquent, une vision unifiée est nécessaire même pour concevoir des Data marts par département.

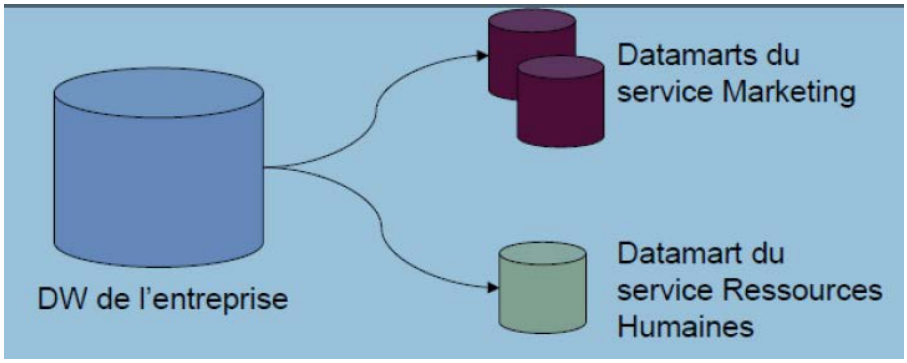


Figure 2.3 Data Marts

IV. Data Mining

IV.1. Définition

«L'Extraction de Connaissances à partir des Données (ECD) est un processus itératif et interactif d'analyse d'un grand ensemble de données brutes afin d'en extraire des connaissances exploitables par un utilisateur analyste qui y joue un rôle central» [12]

A partir de cette définition le Data Mining est l'ensemble des techniques qui permettent de transformer les données en connaissances

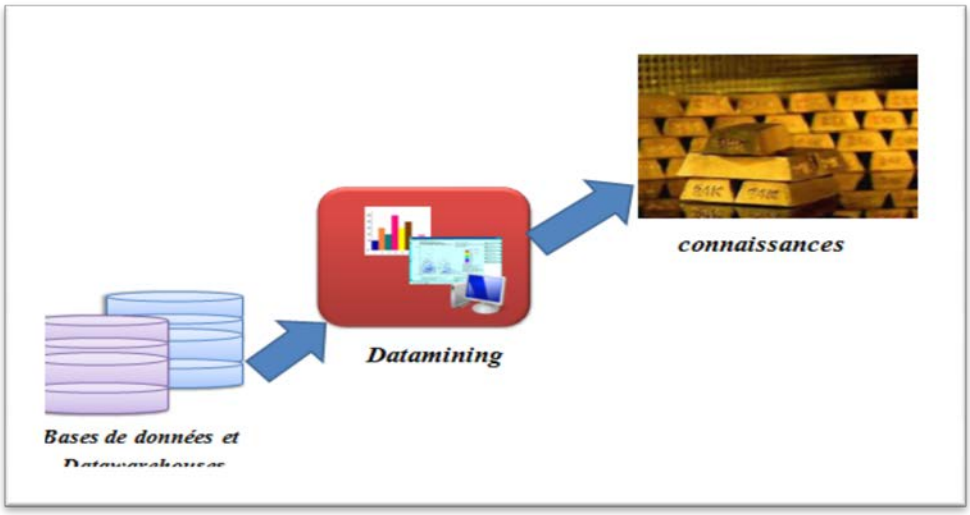


Figure 2.4 Data Mining

IV.2. Principe les méthodes du datamining

Pour arriver à exploiter ces quantités importantes de données, le data mining utilise des méthodes d'apprentissages automatiques. Un amalgame est fait à tort entre toutes ces méthodes. Ces méthodes sont de deux types : les méthodes descriptives et les méthodes prédictives, selon qu'il existe ou non une variable "cible" que l'on cherche à expliquer. [13]

IV.2.1. Méthodes descriptive

Le principe de ces méthodes est de pouvoir mettre en évidence les informations présentes dans le data warehouse mais qui sont masquées par la masse de donnée.
Parmi les techniques et algorithmes utilisés dans l'analyse descriptive, on cite :

- Analyse factorielle (ACP et ACM)
- Méthode des centres mobiles
- Classification hiérarchique
- Classification neuronale (réseau de Kohonen)
- Recherche d'association

VI.2.2. Méthode prédictive

Contrairement à l'analyse descriptive, cette technique fait appels à de l'intelligence artificielle. L'analyse prédictive, est comme son nom l'indique une technique qui va essayer de prévoir une évolution des événements en se basant sur l'exploitation de données stockés dans le data warehouse.

En effet, l'observation et l'historisation des événements peuvent permettre de prédire une suite logique. Le meilleur exemple est celui des prévisions météorologiques qui se base sur des études des évolutions météorologiques passées. En marketing, l'objectif est par exemple de déterminer les profils d'individus présentant une probabilité importante d'achat ou encore de prévoir à partir de quel moment un client deviendra infidèle.
Parmi les techniques et algorithmes utilisés dans l'analyse prédictive, on cite :

- Arbre de décision
- Réseaux de neurones
- Régression linéaire
- Analyse discriminante de Fisher

IV.3. Les taches de data mining

Contrairement aux idées reçues, le data mining n'est pas le remède miracle capable de résoudre toutes les difficultés ou besoins de l'entreprise. Cependant, une multitude de problèmes d'ordre intellectuel, économique ou commercial peuvent être regroupés, dans leur formalisation, dans l'une des tâches suivantes :

- Classification,
- Estimation,
- Prédiction,
- Groupement par similitudes,
- Segmentation (ou clusterisation),
- Description,
- Optimisation.

Afin de lever toute ambiguïté sur des termes qui peuvent paraître similaires, il semble raisonnable de les définir.

V. On-Line Analytical Processing (OLAP)

V.1. Définition

OLAP signifie « **O**n **L**ine **A**nalytical**P**rocessus » repose sur une base de données multidimensionnelle, destinée à exploiter rapidement les dimensions d'une population de données. Le modèle OLAP sera celui du Data Warehouse, il sera construit pour sélectionner et croiser plusieurs données provenant des sources diverses afin d'en tirer une information implicite. Ceci a évolué pour aboutir à une méthode d'analyse permettant aux décideurs un accès rapide et de manière pertinente présentée sous divers angles, dimensions sous forme de cube. L'outil OLAP repose sur la restructuration et le stockage des données dans un format multidimensionnel issues de fichiers plats ou de bases de données relationnelles. Ce format multidimensionnel est connu sous le nom d'hyper cube, ce dernier organise les données le long de dimensions. Ainsi, les utilisateurs analysent les données suivant les axes propres à leur métier. OLAP est un mode de stockage prévu pour l'analyse statistique des données. Une base de données OLAP peut se représenter comme un cube à N dimensions où toutes les intersections sont pré calculées.[14]

V.2. Les différents outils OLAP

- **BIRT :** BIRT est un projet open source basé sur éclipse qui s'intègre à des applications ava/J2EE pour produire des très bons rapports.
- **Pentaho :** Pentaho est une collection de projets open source, principalement axé sur la création, la production et la distribution d'un contenu riche et sophistiqué
- **JasperReports :** JasperReports est un outil de reporting Open source pour le langage Java. Il peut accéder aux données via JDBC, TableModels, JavaBeans, XML, Hibernate, CSV, Il génère des rapports au format PDF, RTF, XML, XLS, CSV, HTML, XHTML, texte, DOCX, et OpenOffice.
- **OpenRPT :** OpenRPT est un outil de reporting multiplateforme (pour Windows, Linux et Mac OS X) qui fait partie de l'ERP xTuple. Il dispose d'une interface écrite en Qt et utilise PostgreSQL comme système de gestion de base de données.
- **OpenReports**: OpenReports est un outil de web reporting puissant souple et facile à utiliser. Il support une variété de moteurs de reporting open source, comme par exemple JasperReports, JFreeReport, JXLS et Eclipse BIRT.
- **DataVision:**DataVision est un outil de reporting Open Source similaire à Crystal Reports. Il dispose d'une interface graphique qui permet de concevoir des rapports en utilisant les glisser-déposer.Il produit des résultats sous différents format comme par exemple HTML, XML, PDF, Excel, LaTeX2e, DocBook.
- **icCube:** est un moteur de traitement analytique multidimensionnel.

disponibles depuis n'importe quel périphérique (ordinateurs, mobiles, tablettes) grâce à son implémentation en Java (normes J2EE).

- **Palo** : Palo est un moteur OLAP et donc une source de données proposant des interfaces : un add-in Excel (Palo for Excel) ainsi qu'une interface web (PaloWorksheet Server).

  Ces outils proposant des fonctionnalités de simulation (saisie/calculs/restitutions) dans le cadre d'une activité de reporting.

V.3. comparaison entre quelques outils OLAP

Jaspersoft iReport, Pentaho Report Designer et Birt sont les trois meilleures offres de reporting Open Source pour cela on essaye de faire une petite comparaison entre eaux :

	Jaspersoft iReport	Pentaho Report Designer	Birt
Stand alone	Oui	Oui	Brique du portail Birt
Performance en génération de rapports	+++	++	+++
Maturité des tableaux croisés dynamiques	+++	+	+++
Intégration des rapports de solutions tierces	+++	+++	+
Reporting mobile	++	++	+
Connexion native cube OLAP	+++	+++	
Couche sémantique des métadonnées	+++	+++	
Facilité de prise en mains	+	+	+++
Possibilités de mise en page	+++	+++	+
Connecteurs natifs Hadoop et NoSQL	+++	+++	
Total	24 / 27	21 / 27	12 / 27

Tab.2.3 comparaison entre quelques outils OLAP

VI. Modélisation multidimensionnelle

VI.1. Niveau conceptuel

Au niveau conceptuel, il existe 2 modèles [15]:

VI.1.1. Modèle en étoile

- Une table de fait centrale et des dimensions
- Les dimensions n'ont pas de liaison entre elles

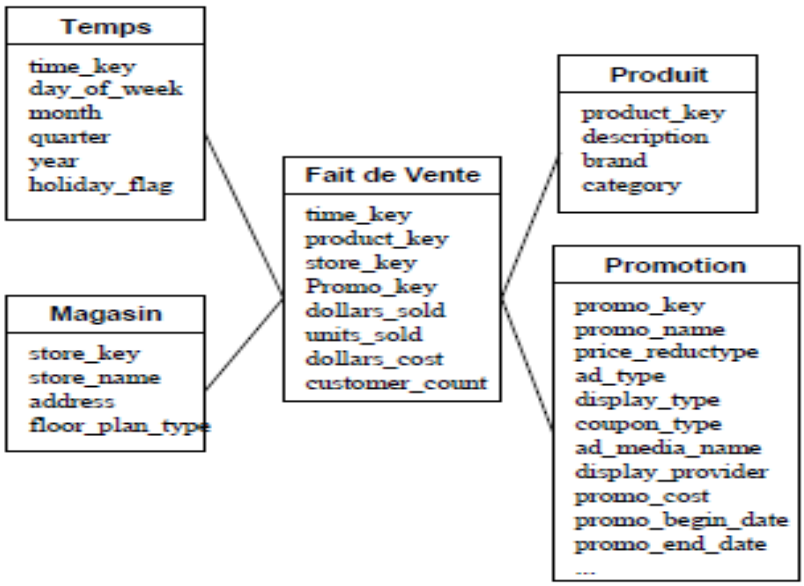


Figure 2.5 Modèle en étoile

VI.1.2.Modèle en constellation (*fact constellation schema*)

Série d'étoiles

Fusion de plusieurs modèles en étoile qui utilisent des dimensions communes

Plusieurs tables de fait et tables de dimensions, éventuellement communes

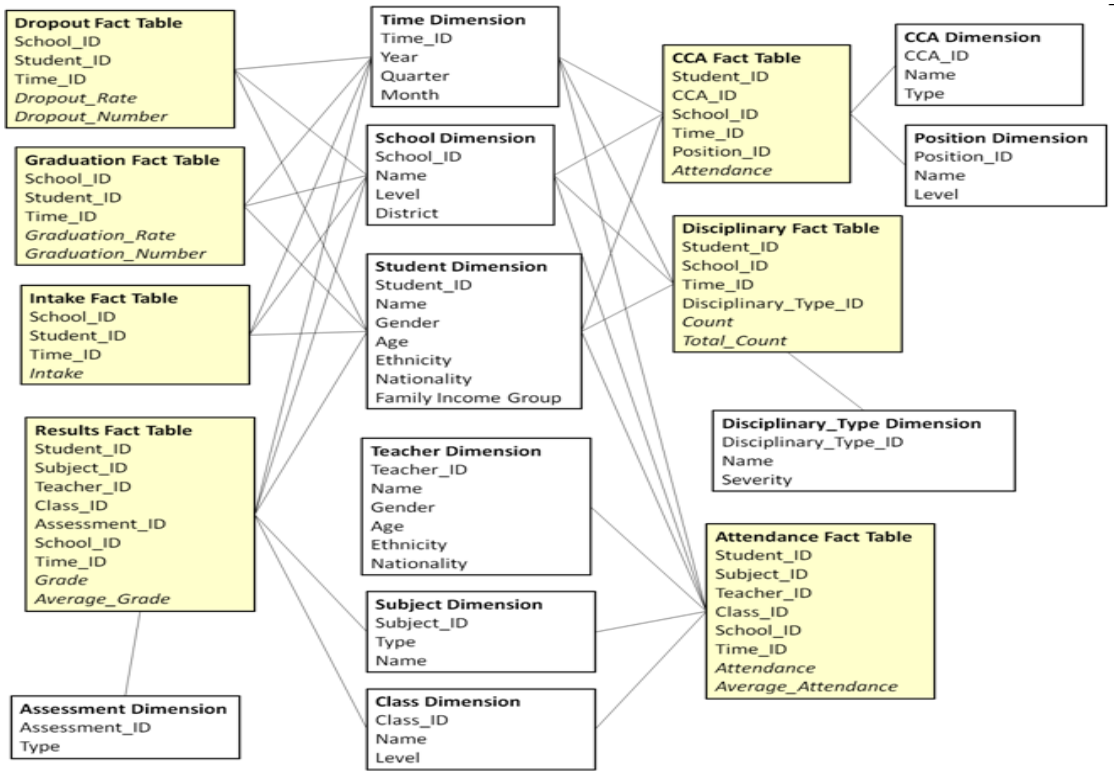


Figure 2.6 Modèle en constellation

VI.2. Niveau logique

Description de la base multidimensionnelle suivant la technologie utilisée :

- ROLAP (*Relational-OLAP*)
- MOLAP (*Multidimensional-OLAP*)
- HOLAP (*Hybrid-OLAP*)

VI.2.1. Multidimensionnel OLAP (MOLAP)

Il est plus facile et plus cher à mettre en place, il est conçu exclusivement pour l'analyse multidimensionnelle avec un mode de stockage optimisé par rapport aux chemins d'accès prédéfinis. MOLAP repose sur un moteur spécialisé, qui stocke les données dans format tabulaire propriétaire (Cube). Pour accéder aux données de ce cube, on ne peut pas utiliser le langage de requête **SQL,** il faut utiliser une API spécifique.

VI.2.2. Relationnel OLAP (ROLAP)

Il est plus facile et moins cher à mettre en place, il est moins performant lors des phases de calculs. En effet, il fait appel à beaucoup de jointure et donc les traitements sont plus conséquents. Il superpose au-dessus des SGBD/R bidimensionnels un modèle qui représente les données dans un format multidimensionnel. ROLAP propose souvent un composant serveur, pour optimiser les performances lors de la navigation dans les données. Il est déconseillé d'accéder en direct à des bases de données de production pour faire des analyses tout simplement pour des raisons des performances.

VI.2.3. Hybride OLAP (HOLAP)

HOLAP est une solution hybride entre les deux (MOLAP et ROLAP) qui recherche un bon compromis au niveau du coût et de la performance. HOLAP désigne les outils d'analyse multidimensionnelle qui récupèrent les données dans de bases relationnelles ou multidimensionnelles, de manière transparente pour l'utilisateur. Ces trois notions se retrouvent surtout lors du développement des solutions. Elles dépendent du software et hardware. Lors de la modélisation, on ne s'intéresse qu'à concevoir une modélisation orientée décisionnelle, indépendamment des outils utilisés ultérieurement.

VI.2.4. Modélisation

Au niveau logique, il existe 1 modèle :

en flocon (*snowflakeschema*)

Modèle en étoile + normalisation des dimensions

- Une table de fait et des dimensions en sous-hiérarchies
- Un seul niveau hiérarchique par table de dimension
- La table de dimension de niveau hiérarchique le plus bas est reliée à la table de fait (elle a la granularité la plus fine)

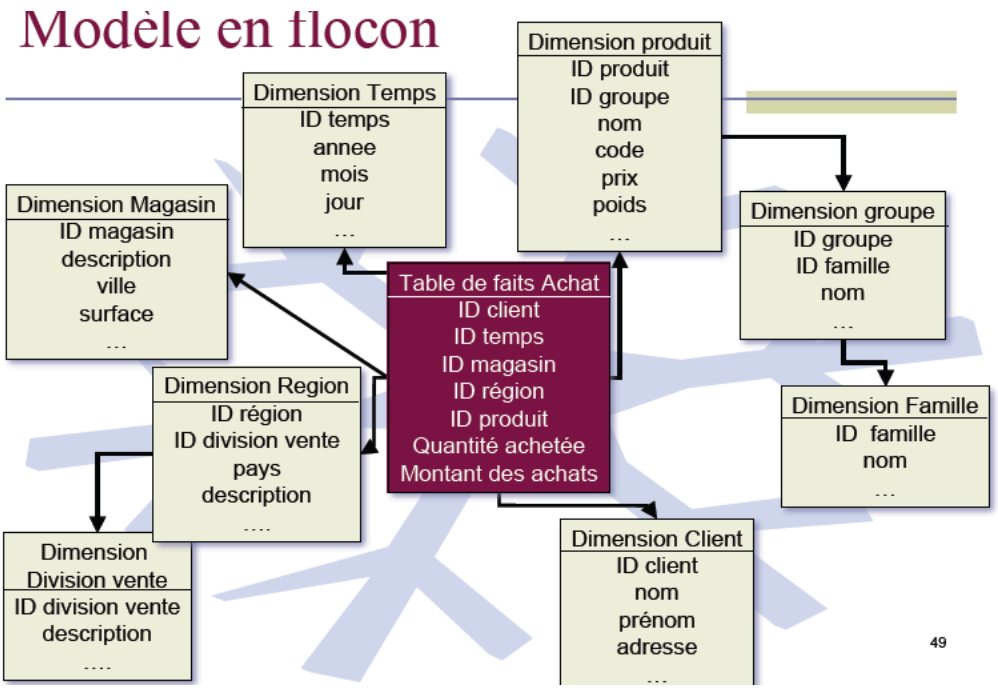


Figure 2.7 Modèle en flocon

VI.3. Niveau physique

- C’est l’implantation et dépend donc du logiciel utilisé.
- Globalement : insuffisance des instructions SQL classiques
- CREATE TABLE … AS … : recopie physique, à reprendre intégralement lors de l’évolution des sources
- CREATE VIEW … AS … : recalculé à chaque requête, temps de réponse inacceptable sur les volumes manipulés
- Optimisation : indexes, …

VII. Synthèse sur les outils pour les entrepôts de données

VII.1. Oracle

Oracle est un SGBD (système de gestion de bases de données) édité par la société du même nom (Oracle Corporation)

La société *Oracle Corporation* a été créée en 1977 par Lawrence Ellison, Bob Miner, et Ed Oates. Elle s'appelle alors *Relational Software Incorporated (RSI)* et commercialise un Système de Gestion de Bases de données relationnelles (SGBDR ou RDBMS pour *RelationalDatabase Management System*) nommé *Oracle*.[16]

VII.2. Hadoop

Il s'agit d'un framework Open Source développé sous l'égide de la fondation Apache, écrit en Java, conçu pour réaliser des traitements sur des volumes de données massifs, de l'ordre de plusieurs petaoctets (soit plusieurs milliers de To). Il s'inscrit donc typiquement sur le terrain du Big Data, nouveau domaine du Cloud Computing.

Hadoop a été conçu par Doug Cutting en 2004. Egalement à l'origine du moteur Open Source Nutch. Doug Cutting cherchait une solution pour accroître la taille de l'index de son moteur. Il eut l'idée de créer un framework de gestion de fichiers distribués.

Yahoo! en est devenu ensuite le principal contributeur. Le portail utilisait notamment l'infrastructure pour supporter son moteur de recherche historique. Comptant plus de 10 000 clusters Linux en 2008, il s'agissait d'une des premières architectures Hadoop digne de ce nom... avant que Yahoo! ne décide de baser son moteur sur Microsoft Bing en

2009. Continuant à recourir à ce socle pour sa gestion de contenu Web et de ses annonces publicitaires, le groupe décide en 2011 de lancer une société (baptisée Hortonworks) avec pour objectif de proposer une offre de services autour d'Hadoop.(URL8)

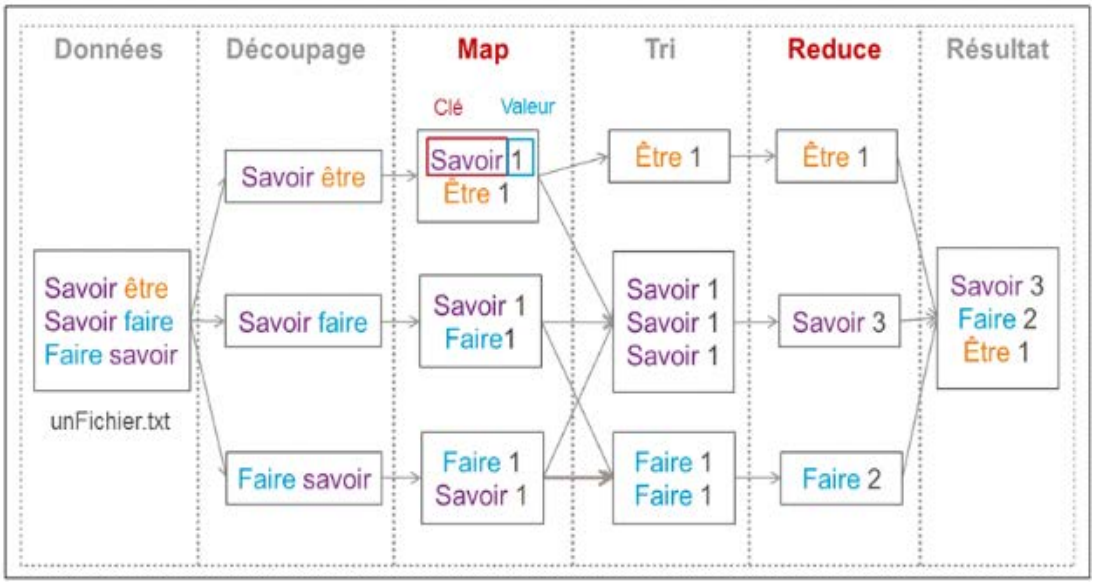


Figure 2.8 MapReduce

VII.3. MySQL

VII.3.1. Définition

MySQL est donc un Système de Gestion de Bases de Données Relationnelles, qui utilise le langage SQL. C'est un des SGBDR les plus utilisés. Sa popularité est due en grande partie au fait qu'il s'agit d'un logiciel Open Source, ce qui signifie que son code source est librement disponible et que quiconque qui en ressent l'envie et/ou le besoin peut modifier MySQL pour l'améliorer ou l'adapter à ses besoins.

Une version gratuite de MySQL est par conséquent disponible. À noter qu'une version commerciale payante existe également.(URL9)

Le logo de MySQL est un dauphin, nommé Sakila suite au concours *Name the dolphin* ("Nommez le dauphin").

VII.3.2. Un peu d'histoire

Le développement de MySQL commence en 1994 par David Axmark et Michael Widenius. EN 1995, la société MySQL AB est fondée par ces deux développeurs, et Allan Larsson. C'est la même année que sort la première version officielle de MySQL. En 2008, MySQL AB est rachetée par la société Sun Microsystems, qui est elle-même

rachetée par Oracle Corporation en 2010.
On craint alors la fin de la gratuité de MySQL, étant donné qu'Oracle Corporation édite un des grands concurrents de MySQL : Oracle Data base, qui est payant (et très cher). Oracle a cependant promis de continuer à développer MySQL et de conserver la double licence GPL (libre) et commerciale jusqu'en 2015 au moins.

Conclusions

Ce chapitre présente les concepts de base de l'entrepôt de données (Data Warehouse) et la différence entre le processus transactionnel en ligne (OLTP) qui gèrent les bases de données représentant le système opérationnel de l'entreprise et le processus d'analyse en ligne de données (OLAP) qui exploite l'entrepôt de données pour les applications d'aide à la décision, on parelle aussi sur le Data Mart qui est un mini Data Warehouse et on fait une petite comparaison entre le Data Mart et le Data Warehouse.
Et en fin les synthèses sur les outils pour les entrepôts de données : Oracle, Hadoop et MySQL.

Chapitre 3

Application

I. Introduction :

Dans ce chapitre en va expliquer les étapes de la réalisation du projet. En a 3 parties, partie 1 la conception d'application on a définis la finalité du DW et aussi on a conçu le modèle de données, partie 2 extraction des donnée ou bien acquisition des données alors dans cette partie on avait besoin de parler d'un outil pour automatiser les chargements de l'entrepôt, et enfin la troisièmes partie et l'implémentation où on va définir les aspects techniques de la réalisation, les modes de restitution et types d'outils de restitution.

II. La Base de données

Pour construire notre entrepôt de donnée il faut passer par les étapes qu'on a vu au chapitre précédent, donc pour cela on commence par l'acquisition des données on utilise EasyPHP DevServer14.1 qui nous permet d'importer les informations et les données de défèrent source, et après pour passer à la deuxième étape : Stockage où les données sont chargées dans une base de données pouvant traiter des applications décisionnelles.

Et enfin la dernière étape : Restitution des données dans cette étapes il existe plusieurs outils mes on a choisi analyse multidimensionnelle.

On a choisi MYSQL car c'est la plus utilisé et aussi moins chère prraport aux autres.

Quelques solutions Open source :

ETL	Entrepôt de données	OLAP	Reporting	Data Mining
▪Octopus ▪Kettle ▪CloverETL ▪Talend	▪MySql ▪Postgresql ▪Greenplum/Bizgres	▪Mondrian ▪Palo	▪Birt ▪Open Report ▪Jasper Report	▪Weka ▪R-Project ▪Orange ▪Xelopes

Table3.1 : solutions Open source

III. Les moyen Utiliser

Pour faire la connexion avec les réseaux sociaux et faire l'analyse de donnée on utilise **:**

- ✓ API Facebook
- ✓ SDK (facebook-php-sdk-3.2.3)
- ✓ API Google+
- ✓ API Twitter
- ✓ MySQL
- ✓ EasyPHP
- ✓ Excel add_in
- ✓ Adobe Dreamweaver CS6

Alors on commence par l'acquisition des données et pour cela on utilise MySQL qui nous permet d'importer les informations et les données de défèrent source.

MySQL d dérive directement de SQL (StructuredQuery Language) est un langage de manipulation de bases de données mis au point dans les années 70 par IBM

_ L'outil phpMyAdmin est développé en PHP et offre une interface pour l'administration de la base de données

_ phpMyAdmin est téléchargeable ici : http://phpmyadmin.sourceforge.net

_ Cet outil permet de :

_ Créer de nouvelles bases

_ Créer/modifier/supprimer des tables

_ Afficher/ajouter/modifier/supprimer des tuples dans des tables

_ Effectué des sauvegardes de la structure et/ou des donnes

_ Effectué des requêtes

_ gérer les privilèges des utilisateurs

L'intérêt de SQL est que c'est un langage de manipulation de bases de données standard permettant de gérer une base de données Access, Paradoxe, dBase, SQL Server, Oracle ou Informix. Une requête SQL prend généralement le format suivant :

SELECT [DISTINCT] attribut(s)

FROM table(s)

[WHERE condition] [GROUP BY field(s)] [HAVING condition] [ORDER BY attribute(s)]

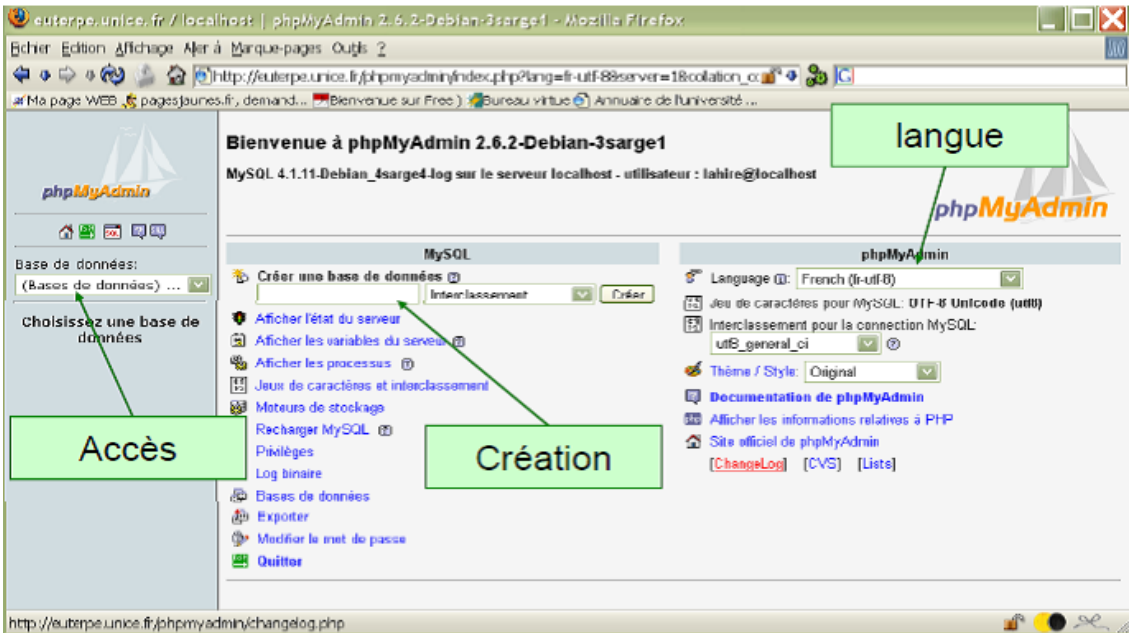


Figure3.1 : L'interface de EasyPHP(URL)

IV. Connexion à la base de données MySQL

Comme nous allons insérer les données du formulaire dans une base de données MySQL, nous allons commencer par nous connecter à celle-ci. Commençons par définir quelques paramètres indispensables pour se connecter : le serveur, le nom d'utilisateur, le mot de passe, et le nom de la base de données.

```
// Parametresmysql à remplacer par les vôtres
define('DB_SERVER','localhost');// serveurmysql
define('DB_SERVER_USERNAME','root');// nom d'utilisateur
define('DB_SERVER_PASSWORD','motdepasse');// mot de passe
define('DB_DATABASE','RS');// nom de la base
```

Sur une ligne en PHP, tout ce qui est précédé par un double slash // est pris comme un commentaire de code, donc vous mettez ce que vous voulez. Remarquez aussi l'utilisation de la fonction define () qui permet de définir une constante (valeur qui ne changera jamais à la différence d'une variable). Une fois ces valeurs constantes définies, vous allez vous connecter à votre serveur de base de données

MySQL en utilisant la fonction mysql_connect () puis choisir une base de données à l'aide de mysql_select_db ().

```
    // Connexion au serveur mysql
    $connect=mysql_connect(DB_SERVER, DB_SERVER_USERNAME,
    DB_SERVER_PASSWORD)
    or die('Impossible de se connecter : '.mysql_error());
       // sélection de la base de données
mysql_select_db(DB_DATABASE,$connect);
```

La ligne commençant par "or" peut vous paraître incompréhensible au premier abord. Si c'est le cas, sachez que le "or" est un "ou logique" et que la fonction mysql_connect () retourne vrai ou faux, suivant si elle réussit ou échoue. Ici si elle retourne "vrai", la ligne du "or" ne se sera pas exécutée. Mais si elle retourne "faux", la ligne sera exécutée et on écrira un message d'erreur avec l'erreur précise qui a eu lieu avec la fonction mysql_error ().

La figure suivant présente l'interface de notre base de donnée elle contient 4tables

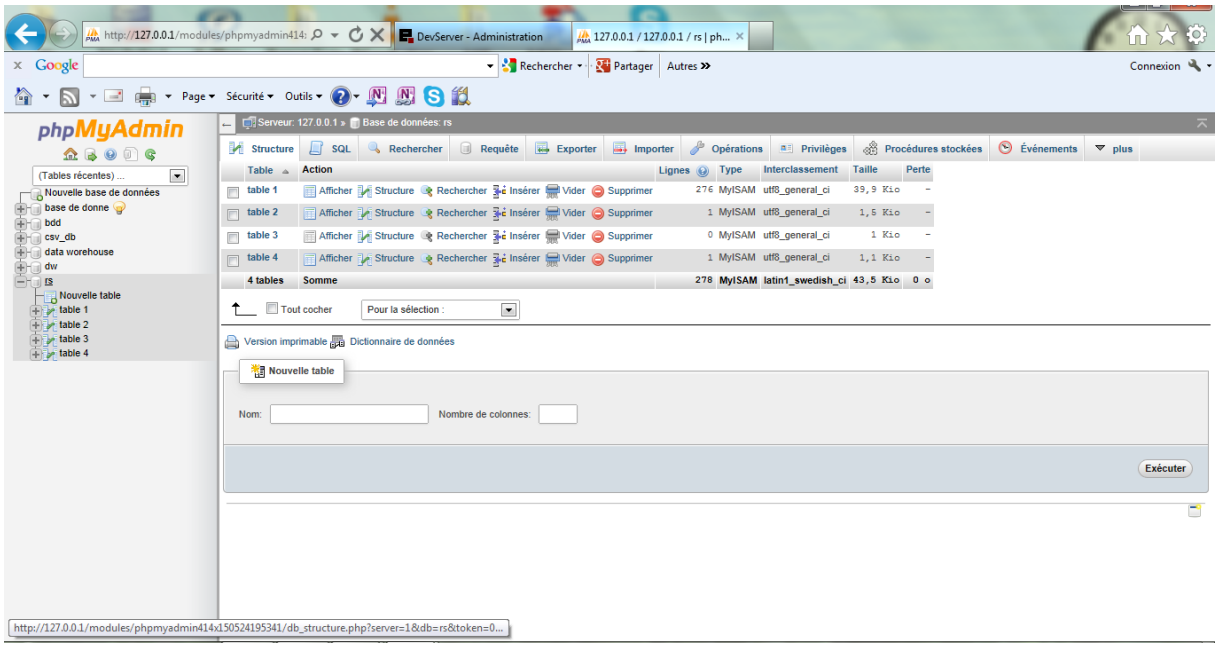
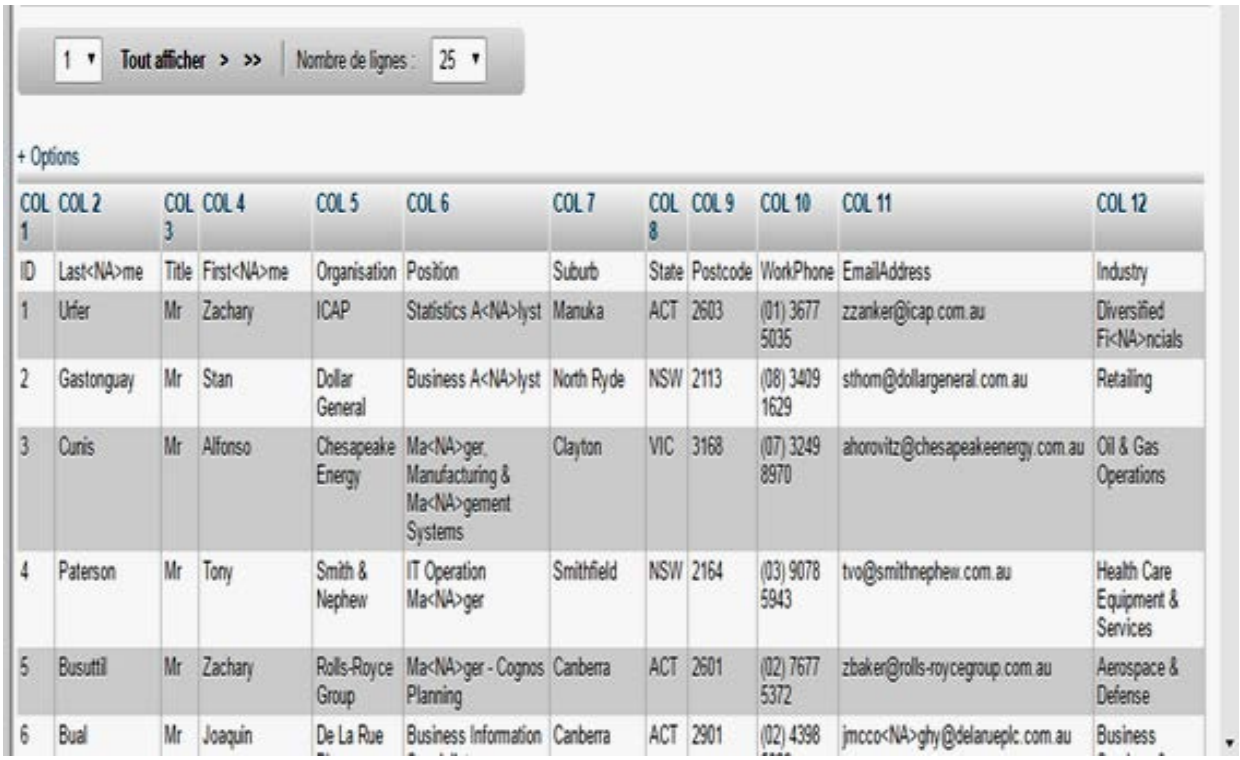


Figure 3.2 : Base de données

Chaque table est contient des informations qui sont télécharger a partir de noter BDD par exemple la première table contienne une liste des personnes, chaque ligne a 12 colonne (nom, prenom , adresse , …)



1 Tout afficher > >> Nombre de lignes : 25

+ Options

COL 1	COL 2	COL 3	COL 4	COL 5	COL 6	COL 7	COL 8	COL 9	COL 10	COL 11	COL 12
ID	Last<NA>me	Title	First<NA>me	Organisation	Position	Suburb	State	Postcode	WorkPhone	EmailAddress	Industry
1	Urfer	Mr	Zachary	ICAP	Statistics A<NA>lyst	Manuka	ACT	2603	(01) 3677 5035	zzanker@icap.com.au	Diversified Fi<NA>ncials
2	Gastonguay	Mr	Stan	Dollar General	Business A<NA>lyst	North Ryde	NSW	2113	(08) 3409 1629	sthom@dollargeneral.com.au	Retailing
3	Cunis	Mr	Alfonso	Chesapeake Energy	Ma<NA>ger, Manufacturing & Ma<NA>gement Systems	Clayton	VIC	3168	(07) 3249 8970	ahorovitz@chesapeakeenergy.com.au	Oil & Gas Operations
4	Paterson	Mr	Tony	Smith & Nephew	IT Operation Ma<NA>ger	Smithfield	NSW	2164	(03) 9078 5943	tvo@smithnephew.com.au	Health Care Equipment & Services
5	Busuttil	Mr	Zachary	Rolls-Royce Group	Ma<NA>ger - Cognos Planning	Canberra	ACT	2601	(02) 7677 5372	zbaker@rolls-roycegroup.com.au	Aerospace & Defense
6	Bual	Mr	Joaquin	De La Rue	Business Information	Canberra	ACT	2901	(02) 4398	jmcco<NA>ghy@delaruepic.com.au	Business

Figure 3.3 : affichage du table1

V. Modèle en étoile

Pour la construction du schéma , La figure 3.3montre le schéma Réseaux sociaux construit. Il est composé de la relation defaits Table de fait et des dimensions Facebook,Twiter,GooglePlus et Temps.

Nous décrivons la structure de chaque relation qui compose le schéma :

- Fait :

Table_fait = {id,id_pfacebook, id_ptwiter, Cle_Temps, id_pgoogl}

- Dimensions :

Dimension facebook ={id_pfacebook,nom,email,sexe}

Dimension twiter ={id_ptwiter,nom,email,langage}

Dimension google+ ={id_pgoogl,nom,email,téléphone}

Dimension temps ={Cle_Temps,année,mois,jour}

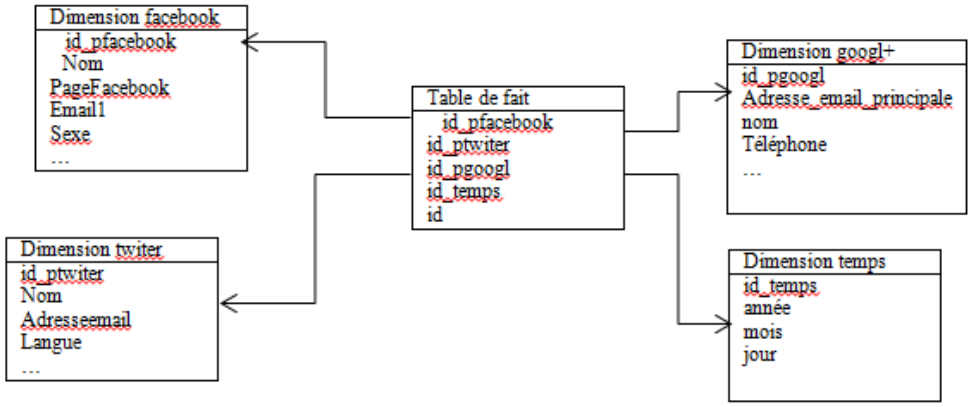


Figure 3.4 : schéma: en étoile

VI. Extraction des schémas locaux à partir des réseaux sociaux

Pour faire l'Extraction des schémas locaux [9]

VI.1 Facebook:

Ce que nous allons faire, c'est se connecter à Facebook depuis notre site, récupérer quelques informations sur notre compte.

Etape1 : L'application Facebook

En crée l'application sur Facebook !

Pour ceci, en un compte Facebook, puis aller sur :

https://developers.facebook.com/apps

En haut il y'a un bouton : <<créer une application>>, cliquez dessus (vraiment besoin de le préciser), un nom pour l'application, une langue, vous acceptez bien sur les conditions d'utilisation de Facebook et vous valider le formulaire.

Etape2 : Téléchargement du SDK et mise en place de l'environnement de travail

Pour cela, en va sur le site github afin de télécharger SDK PHP de Facebook (https://github.com/facebook/php-sdk)

Une foi s votre archive obtenu, naviguez à l'intérieur jusqu'à trouver le dossier src en le désarchive ce dossier, et mettez le à la racine de notre espace de travail.

Etape 3 : On ouvre l'EDI

Créons deux fichiers facebook_actions.php (qui nous servira à faire tous les appels au SDK), avez les bases pour récupérer les informations d'une personne, et même poster sur un mur ! En a exploiter le mode offline.

En remarque que la demande de la permission offline-Access, grâce a cela, l'utilisateur ne doit se connecter qu'une seule et unique fois pour que je puisse utiliser l'API Facebook en son nom.

Fql : Face book Query Langage, ou FQL, vous permet d'utiliser une interface de style SQL pour interroger les données récupérées de Facebook, Le FQL est interprété et traité par l'API Facebook.

VI.2 : Twitter :

L'api Twitter permet d' 'accéder' à la base de données Twitter et de récupérer/poster plusieurs informations. L'API se décompose en quatre classes :

SEARCH : permet d'interroger Twitter pour récupérer des données simples, essentiellement des tweets.

REST : extension avancée de REST qui permet d'accéder à des fonctionnalités avancées de Twitter : chercher des utilisateurs, des followers, voir les status, éditer des informations sur son compte, etc.

STREAMING : permet de communiquer avec Twitter en mode streaming. La particularité de cette api avancée est de permettre l'accès à de gros volume de données Twitter et d'être moins contrainte par les limites d'accès et d'interrogations de Twitter. Cependant, cet api requiert la mise en place de mécanisme plus complexe pour l'accès aux données.

WEBSITES : permet d'intégrer des fonctions de base Twitter dans des sites web.

VI.3 : Google+ :

API g+ : L'API Google+ est l'interface de programmation sur Google+. Cette API permet d'intégrer votre application ou votre site Web à Google+. Elle permet aux utilisateurs de se connecter les uns aux autres à l'aide des fonctionnalités Google+ présentes dans votre application pour susciter un intérêt maximal de leur part.

Remarque :

Les problèmes rencontrés dans la transformation des schémas locaux sont :

-Problème (Interopérabilité)

-les applications change

-problème de permission de l'utilisateur en doit avoir la permission pour récupère le profil.

VI.4. Résultat

VI.4.1 : Schéma locale de Facebook

Voilà le schéma local de Facebook :

Profile Facebook (id_p, Nom, PageFacebook, Email1,Sexe, id, bio, education_school_id1,education_school_n1ame, education_type1, year, education_school_id2, education_school_name2, education_type2, email2, favorite_teams_id1, favorite_teams_name1, favorite_teams_id2, favorite_teams_name2, first_name, gender, hometown_id, hometown_name, last_name, link, location_id, location_name, locale, name, timezone, updated_time, username, verified, work_employer_id, work_employer_name)

Schéma physique de Facebook :

Attributs	Type	Taille
id_p	Int	20
Nom	Varchar	20
PageFacebook	Varchar	20
Email1	Varchar	60
Sexe	Varchar	6
Id	INT	20
Bio	Varchar	20
education_school_id1,	INT	20
education_school_n1ame	Varchar	60
education_type1	Varchar	60
Year	DATE	10
education_school_id2	INT	20
education_school_name2	Varchar	60
education_type2	Varchar	60
email2	Varchar	60
favorite_teams_id1	INT	20
favorite_teams_name1	Varchar	20
favorite_teams_id2	INT	20
favorite_teams_name2	Varchar	60
location_id	INT	2O
hometown_id	INT	2O
hometown_name	Varchar	60
Gender	Varchar	20
first_name	Varchar	2O
last_name	Varchar	20
Link	Varchar	20
location_name, ,	Varchar	60
Locale	Varchar	60
Name	Varchar	20
Timezone	TIME	20
updated_time	DATETIME	20
Username	Varchar	20
Verified	Varchar	20
work_employer_id	Varchar	60

Table3.2 : schéma physique de Facebook

VI.4.2 : Schéma locale de Twitter

Voila le schéma local de Twitter :

profil Twitter (id_p, Nom d'utilisateur, Adresseemail, Langue, Fuseau horaire, Pays, Nom, Localisation, Site Web, Biographie)

Schéma euqisyhp de Twiter:

stubirtta	**epyt**	**elliat**
id_p	INT	20
Nom d'utilisateur	Varchar	20
Adresseemail	Varchar	60
Langue	Varchar	20
Fuseau horaire	TIME	10
Pays	Varchar	20
Nom	Varchar	20
Localisation	Varchar	20
Site Web	Varchar	60

Table3.3 : schéma physique de Twiter

VI.4.3. Schéma locale de Google+

Voila le schéma local de Google+ :

profil Google+ (id_p,Adresse_email_principale, E-mail de récupération de mot de passe, nom, Téléphone, Formation, Sexe, Anniversaire, Situation amoureuse, Autres noms, Autres profils, Liens, Domicile, Professionnel, Adresses, Profession, Compétences, Emplois, langue)

Schéma euqisyhp de Google+:

Attributs	**Type**	**Taille**
id_p	INT	**20**
Adresse_email_principale	Varchar	**60**
E-mail de récupération de mot de passe	Varchar	**60**
Nom	Varchar	**20**
Téléphone	Varchar	**20**
Formation	Varchar	**60**
Sexe	Varchar	**6**
Anniversaire	DATE	**8**
Situation amoureuse	Varchar	**20**
Autres noms	Varchar	**20**
Autres profils	Varchar	**20**
Liens	Varchar	**20**
Domicile	Varchar	**20**
Professionne	Varchar	**60**
Compétences	Varchar	**60**
Emplois	Varchar	**20**
Langue	Varchar	**20**

Table3.4 : schéma physique de Google+

VII. Analyse multidimensionnelle

Le but de l'OLAP (On-Line AnalyticalProcessing) est de permettre une analyse Multidimensionnelle sur des bases de données volumineuses afin de mettre en évidence une analyse particulière des données (il est l'objet d'un questionnement particulier).

Grâce à l'OLAP, les utilisateurs peuvent créer des représentations multidimensionnelles (appelées hypercubes ou « cubes OLAP ») selon les critères qu'ils définissent afin de simuler des situations.

1. Création du cube Olap

On a choisis Palo for ExcelPalo est un moteur OLAP.

Figure3.5 :L'outil Palo

2. A propos de Palo OLAP

Palo OLAP est un logiciel de base de données multidimensionnelle qui se targue de vitesse. Le logiciel en mémoire OLAP a la capacité d'être 100 fois plus rapide qu'une base de données sur disque, ce qui lui permet de soutenir formules multidimensionnelles et données reprises. En outre, ce produit fonctionne en temps réel pour produire des calculatrices et des modèles. Palo OLAP est open source et peut être personnalisé grâce à des interfaces série en PHP, C ++, Java, .NET et VBA.

Palo est possédée par Jedox AG et fournit une suite de logiciels de BI open source. La société a été en affaires depuis 2007 et offre en plus de soutien et de formation des programmes mondiaux.

3. Palo OLAP Caractéristiques principales

•Logiciel de base de données open source gratuit multidimensionnelle qui fonctionne en mémoire

•Jusqu'à 100 fois plus rapide qu'une base de données sur disque

•Peut effectuer données reprises et soutient formules multidimensionnelles

•Personnalisable via PHP, C ++, Java, .NET et VBA

•Permet de gérer les droits des utilisateurs

•Prend en charge jusqu'à 256 dimensions de données

•Calculs en temps réel et la modélisation

Apres la création de notre base de donné nommer RS avec phpmy admin en fait l'importation des donnes vers Exceladd in qui nous permis d'utiliser notre outil d'analyse palo .

Donc pour réaliser notre but de ce projet qui est l'analyse des données qui se base sur un champ par exemple l'Age ou bien adresse…etc., on passe par les étapes suivantes :

On commence par la création de base de données, pour cela on utilise Palo Excel Add-inon clique sur palo et après on appuie sur outil de modélisation

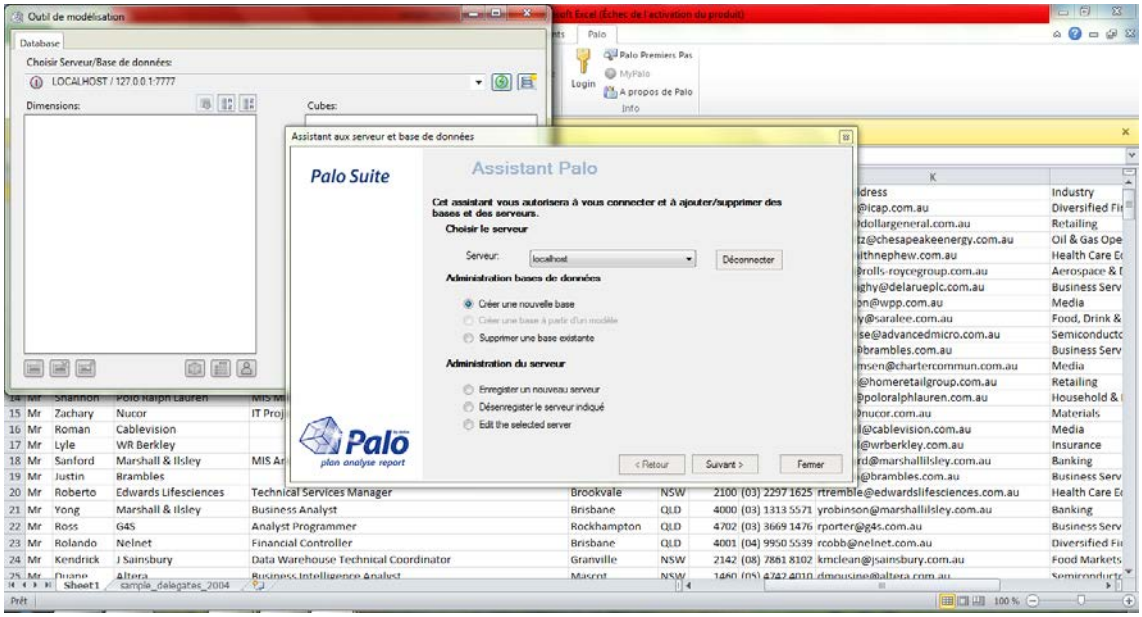


Figure3.6 : Création de BDD

La figure suivante présente l'étape de création de cube et leurs éléments

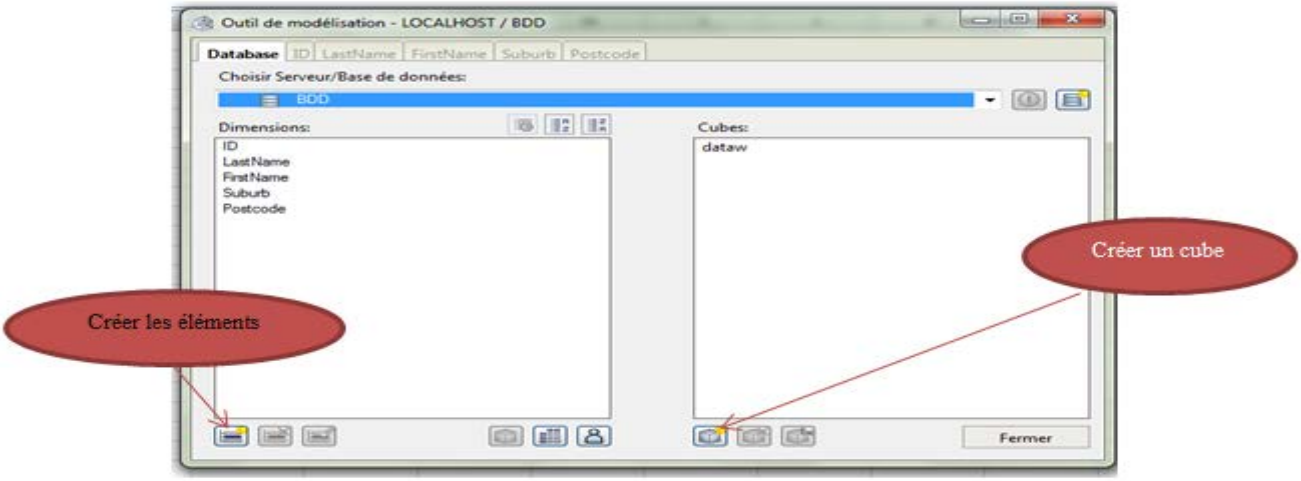


Figure3.7 : créations de cube

Le fonctionnement de notre cube basé sur la création des vues, dans cette étape on choisit les axes et on colle la vues.

C'est quoi une vue dans BDD ?

> Une vue V est une relation qui contient le résultat d'une requête Q évaluée sur une base de données.

Par exemple :

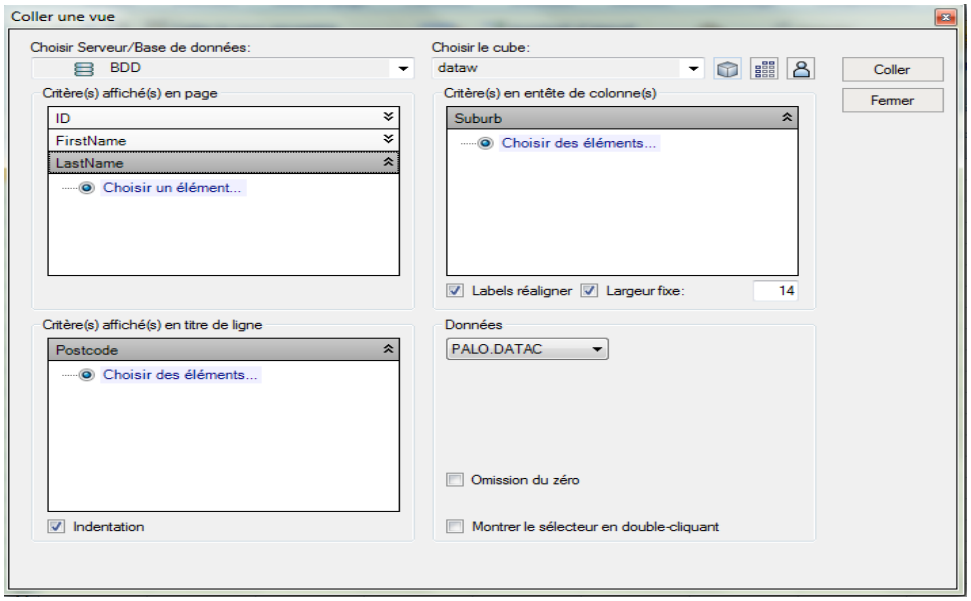


Figure3.8 : colles la vues

Le diagramme suivant présente le résultat d'analyse et de filtre les données et les informations concernant les personne inscrit dans les trois réseaux sociaux Facebook twitter et Google+ de notre base de donnée RS selon les axe suivants : ID, date de naissance, sexe et domaine de travaille.

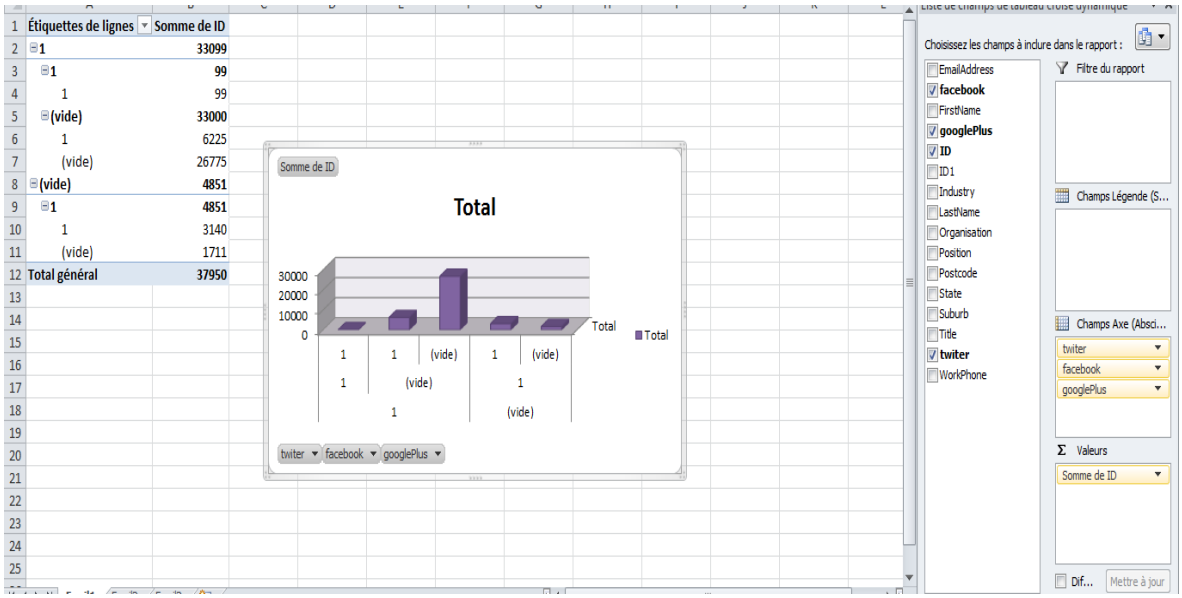


Figure3.9: analyse de donnée

VIII. Conclusion

Dans ce chapitre c'est la partie application, Nous avons utilisé 3 réseaux sociaux : Facebook, Twitter et Google+ pour collecter les données de notre entrepôt de données pour cela on utilise MySql comme SGBD et PHP comme langage. Après la construction de ED on passe l'analyse de données, dans cette étapes on essaye de créer le cube Olap avec l'outil Paloqui est possédée par Jedox.

Conclusion Général

Le but de ce projet c'est la construction d'un entrepôt de donnée pour une identité sociale pour les réutilisations par différents réseaux sociaux. Mais Jusqu'à présent peu de travaux ont considère les problèmes de modélisation et d'implantation d'entrepôt de donnée.

Dans le premier chapitre nous avons présente quelque définitions sur le web 2.0 et les réseaux sociaux puis on a présenté les réseaux connus (Facebook, Twitter, Google plus).

Dans le second chapitre, nous avons présenté des notions et des concepts sur les entrepôts de données, décrivant les Concepts principaux des entrepôts de donnée data mart et data mining et les différents outils OLAP.

Enfin, dans le troisième chapitre, nous avons présenté les outils utilisés pour la réalisation de notre entrepôt de données, et l'explication des étapes de modélisations et d'Extraction des schémas locaux à partir des réseaux sociaux et de création du cube olap.

Nous admettons que par manque de temps, nous n'avons pas pu optimiser toutes les fonctions proposées dans le cadre de ce travail comme l'analyse de donnée avec le data mining et aussi on a eu plusieurs difficultés, Il y a plusieurs réseaux sociaux alors il faut que les données sont à jour.

Les données des réseaux sociaux par fois n'est pas valide c.-à-d. ou ne sont pas vrais ou ne sont pas complets. Donc Les recherches futures éventuelles pourraient s'orienter vers:

-Développer cette application on ajouton la synchronisation des deux sites.

-Intégrer cette application dans d'autres réseaux.

-Généraliser l'application pour les autres réseaux sociaux.

Bibliographie

[1] MarrderNinozka, Facebook et la réalité des amis virtuels, l'Université de Montréal, octobre 2008.

[2]Hugo Lauras, L'impact des réseaux sociaux sur les entreprises a-t-il un rôle essentiel sur leur image.

[3]Antoine Crochet**.**nouvelle pépite montante du Big Data *visiter*12/06/2015

[4] Erick Stattner « Introduction à l'Analyse des Réseaux Sociaux » Université des Antilles et de la Guyane, Guadeloupe, France, Novembre 2012

[5]Agence social media,Comprendre les réseaux sociaux,visiter le *10/06/2015.*

[6]vortexsolution.Réseaux sociaux - Les réseaux sociaux populaires sur le Web a

Montréal.com le 10/09/2015

[7] Françoise Soulié et Emmanuel Viennet ,Réseaux sociaux, analyse et data mining,ÉCOLE NORMALE SUPÉRIEURE PARIS, mars 2011.

[8] http// : www.businessdecision-eolas.com visiter le 08/09/2015.

[9] :www.proximamobile.fr/article/les-perspectives-des-reseaux-sociaux visiterle *10/06/2015.*

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

[11] Séraphin LOHAMBA OMATOKO « Analyse et détection de l'attrition dans une entreprise de télécommunication », Université Notre Dame du Kasayi - Licencié en sciences informatique/Génie Logiciel 2011 .

[12]D. A. Zighed, Y. Kodratoff, and A Napoli, "Extraction de connaissance à partir d'une base de donnée," Bulletin AFIA'01, 2001.

[13] *Bernard CLÉMENT «* Introduction au Data Mining » *Montréal, Canada., 2013*

[14] A. Doucet and S. Gangarski. Entrepôts de données et Bases de Données Multidimensionnelles, Chapitre 12 du livre : Bases de Données et Internet, Modèles, langages et systèmes. Editions Hermès, 2001.
[15] DIB Sidi Mohammed Fazil. LARABI Nor El Islam, Construction d'une identité sociale fédérée représenté en XML par une approche d'intégration de données, Université Abou BakrBelkaid– Tlemcen, 2014.

[16]http://www.oracle.com le 15/04/2015

[17**]** http://www.Présentation de Hadoop - opentuto.com le 08/06/2015
[18]http://www.*David Axmark, fondateur de MySQL*.com le 08/06/2015

.

Liste des Figures

Liste des tableaux

Résumé :
Le travail présenté dans ce mémoire concerne la construction d'un entrepôt de donnée d'identité social. Avec l'avènement du Web 2.0, les plates-formes de réseaux sociaux numériques (RSN) (Facebook, Twitter, GooglePlus, etc.) fournissent en plus de la structuredes réseaux, beaucoup d'informations propres aux individus et à leurs interactions au moyen d'applications multiples et variées (photos, vidéos, tags, blogs, murs,liens, etc.). Plusieurs techniques ont également été développées pour extraire des informations sur les usages des internautes dans ces environnements.La construction d'un entrepôt d'identité social est une tâche importante pour les individus et le producteur d'informations. Les nombreuses données sont analysée et filtrer par l'outil et palo de jedox pour bien organise l'entrepôt de donnée et la recherche d'information sera facile a trouver un résultat pertinent dans un temps court. Pour réaliser ce travail on a utiliséMySql comme SGBD, On choisit le PHP comme langage de programmation. Tout ça pour faciliter ou améliorer plusieurs tâches telles que la détection de communautés basée sur des centres d'intérêts, la détection de réseaux d'influences, la prise en compte du réseau social d'un utilisateur dans la conception de son profil, etc.
Mots clé :
entrepôt de donnée,Web 2.0,réseaux sociaux, MySql, PHP, l'outil palo
Abstract
The work presented in this thesis concerns the construction of a data warehouse of social identity. With the advent of Web 2.0 platforms of digital social networks (RSN) (Facebook, Twitter, GooglePlus, etc.) provide in addition to the structure of the networks, a lot of information specific to individuals and their interactions through multiple and varied applications (photos, videos, tags, blogs, walls, links, etc.). Several techniques have also been developed to extract information from the uses of Internet users in these environments. The construction of a social identity warehouse is an important task for individuals and producers of information. Many data are analyzed and filtered by the tool Jedox Palo to properly organize the data warehouse and the search for information for facilitate the finding of relevant results quickly. To carry out this work we have used MySql as DBMS,and the PHP programming language. All that to facilitate and improving several tasks such as sensing communities based on interests, detection of influence networks, taking account of the social network of a user in the design of its profile, etc.

Key words:

data warehouse, Web 2.0, social networks, MySql, PHP, the palo tool

ملخص:

(RSN) العملالمقدمفيهذهالأطروحةيتعلقبناءالهويةالاجتماعيةمستودعالبيانات. معظهورالويب 2.0 منصاتالشبكاتالاجتماعيةالرقمية ،الخ)،بالإضافةإلىتوفربنيةالشبكات،والكثيرمنالمعلوماتالخاصةللأفرادوتفاعلاتهامنخلالتطبيقاتمتعددةومتنوعة GooglePlusالفيسبوك،تويتر، (الصوروأشرطةالفيديو،والعلامات،بلوق،والجدران،وصلات،وماإلىذلك).
كماتمتطويرالعديدمنالتقنياتلاستخراجمعلوماتحولاستخداماتمستخدميالانترنتفيهذهالبيئات.
بناءمستودعهويةالاجتماعيهومهمةهامةللأفرادوالمعلوماتالمنتجة.
Jedoxيتمتحليلالعديدمنالبياناتوتصفيتهامنقبلأداةو
بالولتنظيمصحيحمستودعالبياناتوسوفالبحثعنالمعلوماتيكونمنالسهلالعثورعلىنتيجةذاتالصلةفيوقتقصير.
PHP. لتنفيذهذاالعملاستخدمنالخليةكمانظمإدارةقواعدالبيانات،هواختيارلغةالبرمجة
كذلكلتسهيلأوتحسينالعديدمنالمهاممثلاالمجتمعاتالاستشعارتقومعلىالمصالح،والكشفعنشبكاتالنفوذ،معالأخذفيالاعتبارالشبكةالاجتماعيةللمستخدم فيتصميمصورتها،الخ

الكلماتالرئيسية:

أداةpalo،PHP,مستودعالبيانات،الويب 2.0،والشبكاتالاجتماعية،