

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université de Larbi Tébessi-Tébessa-

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

Département : Mathématiques et Informatique



MEMOIRE DE MASTER

Domaine : Mathématiques et Informatique

Filière : Informatique

Option : Système d'information

Thème

Intégration du Big-Data et la Médecine

Personnalisée

Présenté par

- Warda Loucif

- Bariza Belghit

Devant le jury:

Y. Mennassel	MAA	Université Laarbi Tébessi: Président
M.C. Nait Hamoud	MAA	Université Laarbi Tébessi: Examineur
Louardi Bradji	MCB	Université Laarbi Tébessi: Encadreur

Année Universitaire: 2016-2017

ملخص:

تلعب التكنولوجيا دورا كبيرا في أيامنا هذه خاصة في مجال الإعلام الآلي والمعلوماتية.

المعلومات التي يتم جمعها ودراستها وحفظها تسمى البيانات الكبيرة.

يعتمد الطب الشخصي أساسا على استعمال المعلومة الوراثية والجزيئية.

في دراستنا هذه قدمنا نموذج تكامل البيانات الكبيرة مع عمليات استخراج المعلومات وتحويلها وتحميلها مع خدمات الويب، اعتمادا على الأعمال ذات صلة وحاولنا حل عدم التجانس باستخدام خدمات ويب ولغة الترميز الموسعة.

كلمات مفتاحية:

البيانات الكبيرة، الطب الشخصي، طريقة ETL ، إدماج المعطيات الغير متناسقة

Résumé :

La technologie joue un grand rôle aujourd'hui, en particulier dans le domaine de l'information automatisée et informatique.

Les informations recueillies et étudiées et enregistrées sont appelées Big Data.

La médecine personnalisée est une action nouvelle ayant pour principe d'utiliser les informations génétiques on moléculaires.

Notre étude, nous avons fait ces grandes intégration de données avec un ETC (extraction, transformation et chargement) modèle pratique avec les services Web, en fonction de l'activité concernée et nous avons essayé de résoudre l'hétérogénéité en utilisant des services Web et XML.

Mots-clef : Big Data, Médecine personnalisée, approche ETC, intégration de données hétérogènes.

Abstract:

Technology plays a big role today, especially in the field of computing and informatics.

The information collected, studied and saved is called big data.

Personalized medicine relies mainly on the use of genetic and molecular information.

In our study we presented big data integration model with the operations of extracting information, transforming and uploading them with web services, depending on the relevant work and we attempted to resolve the heterogeneity using Web services and extended coding language.

Key-words: Big Data, Personalized Medicine, ETL approach, heterogeneous data integration.



Remerciements

Au terme de la rédaction de ce mémoire, nous tenons à remercier notre encadreur Mr.Dr; Bradji Louardi .

A pour leurs précieux conseils et leur aide durant toute la période du travail.

Nos vifs remerciements vont également aux membres du jury pour l'intérêt qu'ils ont porté à notre recherche en acceptant d'examiner notre travail Et de l'enrichir par leurs propositions .

Merci



Sommaire :

Titre	Page
Résumé	I
Liste des figures	VI
Liste des tableaux	VII
Introduction générale	01
Chapitre 01: Aperçu sur le Big Data	
1. Introduction	04
2. Big Data	05
3. Caractéristiques des Big Data	05
3.1. Volume	05
3.2. Vitesse	06
3.3. Variété	06
4. Les nouveaux enjeux et risques du Big Data	06
4.1. Enjeux techniques	07
4.2. Enjeux économiques	07
4.3. Enjeux juridiques	08
5. L'analyse : le point clé du Big Data	08
6. Processus de chargement et de collecte de données dans big data	08
7. Technologie des Big Data	10
7.1. Apache HADOOP	10
7.2. Modèle Mapreduce	10
8. Data warehouse	12
9. Les caractéristiques de données d'entrepôts	13
10. Composantes d'un entrepôt de données	13
10.1. Les applications opérationnelles sources	13
10.2. La préparation des données	13
10.3. Les outils d'accès aux données	14
11. Les données dans l'entrepôt	14
12. Principes d'un entrepôt de données	14
12.1. En amont et en aval (« front-end » et « back-end »)	14
12.2. Dimensions et faits	15
12.3. Architecture d'un entrepôt	15
12.4. Serveurs OLAP(On-Line Analytical Processing)	15
13. Construction d'un entrepôt de données	16
13.1. La planification	16
13.2. Conception et implémentation	16
14. L'intégration de données	17
14.1. Les composants d'un système d'intégration	17
14.2. Caractéristique des sources	18
15. Approches d'intégration de l'entrepôt de données	18
15.1. Entreprise information intégration (EII)	18

15.2. Entreprise application intégration (EAI)	19
15.3. Extract, Transform and Load (ETL)	20
Conclusion	22
Chapitre 02: Informatique médicale et Médecine personnalisée	
1. Introduction	24
2. Informatique médicale	25
2.1. Définition	25
2.2. Les Grands domaines de l'informatique médicale	26
2.2.1. L'EBM (Evidence Based Medicine)	26
2.2.2 La médecine personnalisée	26
2.3. Importance de l'informatique dans le secteur de la santé	27
2.3.1. La disponibilité de l'information	27
2.3.2. Gestion de base de données	27
2.3.3. Technologie avancée	28
2.3.4. Imagerie médicale	28
2.4. Principales applications de l'informatique médicale	28
2.4.1. Systèmes documentaires et banques d'information	29
2.4.2. Le dossier médical informatisé (DMI)	29
2.4.3. Systèmes d'aide à la décision médicale	30
3. Médecine personnalisée	31
3.1. Définitions	31
4. Big data et santé	33
4.1 Big Data et médecine personnalisée	33
5. Système d'aide à la décision médicale	34
5.1. Décision médicale	34
5.2. Définition d'un système d'aide à la décision médicale	34
5.3. Différents types de systèmes d'aide à la décision médicale	35
5.3.1. les systèmes d'aide indirecte à la prise de décisions ou d'assistance documentaire	35
5.3.2. les systèmes d'alerte ou de rappels automatique	35
5.3.3. les système consultants ou " Aide centrée patient " (patient – centred support)	35
6. Les systèmes experts pour l'aide à la décision médicale	35
1.1. Définition	36
1.2. Architecture d'un système Expert	36
6..2.1. composant du système expert	37
Conclusion	38
Chapitre 03: Contribution, modèle d'intégration avec ETL et service web	
1. Introduction	40
2. Les usages d'entrepôt de données dans la Médecine	41
2.1. L'intégrer des sources des données patientes hétérogènes	41
3. Intégration des données de prescription dans un entrepot de données	42
4. Entreposage de données complexes pour la médecine personnalisée	43
5. Notre proposition	44
5.1. Sources Hétérogènes	45

5.2. Service Web	45
5.3. Data WareHouse	45
5.4. OLAP	45
6. L'implémentation de modèle	46
6.1. Technologies XML	47
6.2. Sources hétérogènes	48
6.3. Service Web et ETL	49
Conclusion	54
Conclusion générale	55
Bibliographie	56

Liste des figures :

Titre	Pages
Figure 1.1: Couche de chargement des données dans le Big Data	09
Figure 1.2: Exemple d'un programme MapReduce (WordCount)	12
Figure 1.3: Composantes de base détaillées d'un entrepôt de données	14
Figure 1.4: Architecture générale d'un système d'intégration	17
Figure 1.5: Approches d'intégration de l'entrepôt de données	18
Figure 2.1: Architecture de système expert	37
Figure 3.1 : données cliniques hétérogènes	41
Figure 3.2 : Schéma XML d'un élément de prescription	42
Figure 3.3 : Architecture globale de l'entrepôt de données	43
Figure 3.4 : système d'informatique hospitalier	44
Figure 3.5 : Visual Studio	46
Figure 3.6 : Résultats du labo type Excel	48
Figure 3.7 : liste de préinscription type Access	48
Figure 3.8 : liste des consultations type SQL	48
Figure 3.9 : Interface d'application d'extraction	59
Figure3.10 : Interface d'application d'extraction en exécution	59
Figure 3.11 : l'interface de ETL au niveau de service web	50
Figure 3.12 : la méthode ETL au niveau de service web	50
Figure 3.13 : connexion entre l'application d'extraction et service web	51
Figure 3.14 : la sélection des champs et sources	52
Figure 3.15 : l'appelle de méthode ETL au niveau d'application d'extraction	52
Figure3.16 : le Data Warehouse	53

Liste Des Tableaux :

Titre	Pages
Tableau 1.1 : Alternatives de mise en œuvre OLAP	16
Tableau 1.2 : Comparaison entre les approches d'intégration	21

Introduction Générale

Actuellement, la pratique médicale a fait des pas de géant grâce à des techniques qui ont révolutionné ce domaine sensible : désormais, la tâche du praticien n'est plus figée comme auparavant. Au contraire, il doit accomplir énormément d'opérations et adopter d'autres comportements au service du patient. Il doit même rendre des comptes à ses patients, sur tous ses agissements.

Le médecin est tributaire de ses patients, qui deviennent de plus en plus exigeants quant à une prise en charge efficace. Aujourd'hui, les patients se sont émancipés car ils sont conscients de leur état et des complications de leurs maladies.

La crédibilité de la communauté médicale est étroitement liée aux agissements de tous les confrères, autrement dit de tout un chacun car il y va de la survie d'une étique d'une profession. En effet, une faute médicale se répercute sur l'ensemble des médecins qu'ils soient concernés de près ou de loin.

Les avantages de l'outil informatique dans le domaine médical est très important, cependant il auxquels on n'accorde pas beaucoup d'importance et qui pourtant peuvent avoir des influences décisives de pratique ou de modification.

La « technique » au sens de la méthode (basée sur le recueil des symptômes) permettant de donner un diagnostic et de proposer un traitement est retenu parmi les différents aspects de la « technique médicale ». « L'informatique médicale » a de beaux jours devant lui, car les prouesses médicales en dépendent. On peut dire que l'informatique a révolutionné la médecine moderne.

La question qui s'impose est: l'informatique ne risque –t– elle pas de porter atteinte à notre réceptivité à certaines données cliniques?

Dans un souci d'efficacité, de modernisation et d'économie, nous adhérons pour une recherche de solutions ad hoc en vue de gérer la complexité des connaissances scientifiques et la mise en œuvre des procédures et des systèmes d'information adéquats à la pratique médicale.

Notre objectif est de présenter une vision globale de le Big data et médecine personnalisée, Nous avons également essayé de résoudre le problème d'hétérogénéité en utilisant les services Web et XML.

Ce document s'articule en quatre (3) chapitres:

Chapitre 1 : Dans ce premier chapitre nous allons définir -en premier lieu- du big data . En deuxième lieu nous allons monter l'importance de Big data ; puis les caractéristiques de Big data entre autre le volume, la vitesse et la variété ; aussi nous allons parler des enjeux et risques de Big data et surtout nous allons entamer les entrepôts des données et leur intégration hétérogène .

Chapitre 2 : ce chapitre sera consacré à la description de l'informatique médicale et leur Importance dans le secteur de la santé. Et ses grands domaines notamment la médecine personnalisée. Dans ce chapitre nous allons parler aussi du système d'aide à la décision médicale et le système expert .Nous allons définir le système expert et de son architecture et ses composants.

Chapitre 3 : Dans ce dernier chapitre nous allons présenter les usages d'entrepôt de données dans la Médecine personnalisée et de monter notre proposition qui se base sur les travaux précédents et de monter l'objectif principal de notre travail qui consiste dans la proposition et l'implémentation d'un modèle pour les SIH capable d'intégrer plusieurs sources hétérogènes au but de construire un entrepôt de données pour les informations des patients.



Chapitre 01:

Aperçu Sur Le Big Data

1. Introduction
2. Big Data
3. Caractéristiques des Big Data
 - 3.1. Volume
 - 3.2. Vitesse
 - 3.3. Variété
4. Les nouveaux enjeux et risques du Big Data
 - 4.1. Enjeux techniques
 - 4.2. Enjeux économiques
 - 4.3. Enjeux juridiques
5. L'analyse : le point clé du Big Data
6. Processus de chargement et de collecte de données dans big data
7. Technologie des Big Data
 - 7.1. Apache HADOOP
 - 7.2. L'algorithme Mapreduce
 - 7.2.1. Définition
8. Data Warehouse
9. Les caractéristiques de données d'entrepôts
10. Composantes d'un entrepot de données
 - 10.1. Les applications opérationnelles sources
 - 10.2. La préparation des données
 - 10.3. Les outils d'accès aux données
11. Les données dans l'entrepôt
12. Principes d'un entrepot de données
 - 12.1. En amont et en aval (« front-end » et « back-end »)
 - 12.2. Dimensions et faits
 - 12.3. Architecture d'un entrepot
 - 12.4. Serveurs OLAP (On-Line Analytical Processing)
13. Construction d'un entrepot de données
 - 13.1. La planification
 - 13.2. Conception et implémentation
14. L'intégration de données
 - 14.1. Les composants d'un système d'intégration
 - 14.2. Caractéristique des sources
15. Approches d'intégration de l'entrepôt de données
 - 15.1. Entreprise information intégration (EII)
 - 15.2. Entreprise application intégration (EAI)
 - 15.3. Extract, Transform and Load (ETL)

Conclusion

1. Introduction

Les systèmes d'information de la prise de décision repose sur l'entrepôt de données et d'analyse sur le web (*On-Line Analytical Processing* - OLAP) et sont basés sur l'informatique BI (Business Intelligence - BI), est aujourd'hui confrontée à de nouveaux défis scientifiques et technologiques, et basé sur l'intelligence de l'entreprise traditionnelle en fait sur l'entrepôt de données, qui facilement échangé des dizaines de tailles (même des centaines) de téraoctets peu de données ou pas structurées (texte, graphiques et données continues) et l'analyse en temps réel. BI traditionnelle doit évoluer pour tenir compte des nouveaux défis.

Avec une grande apparence en taille et variation de vitesse des technologies émergentes sur les grandes méga-données (ou les grandes données le plus souvent désignées sous le nom de Big Data) et avec l'extension des algorithmes d'exploration de données.

En fait, la propagation de la friture ou pas de flux ordonnée (données provenant de capteurs ou médias sociaux), qui produit une quantité de BI très importantes et inattendues, mais l'entrepôt de données et des analyses sur l'Internet embrassent de nouveaux domaines.

En raison de son rôle clé dans la gestion des organisations, tout en maintenant sa force et comment soutenir l'entreposage des données et des analyses sur l'Internet dans les grands développements dans les données pour les problèmes que nous les trouver ont toujours une place dans les entreprises pour organiser et analyser les nouvelles données historiques.

Ce développement soulève de nouvelles questions scientifiques et technologiques qui nécessitent l'approche de la définition de l'architecture, l'intégration, la requête, l'analyse, l'optimisation, la modélisation et la sécurité dans les entrepôts de données, que ce soit des données de données traditionnelles ou à grande échelle. Ce Big data nous conduit à l'ère des nouveaux bâtiments de la période d'abonnement scientifique et technologique et de l'infrastructure (Hadoop, Nosql,...).

En effet ce chapitre sera consacré en particulier au Big data et l'entrepôts des données et leurs caractéristiques et qu'est-ce qu'un processus d'intégration des données.

2. Big Data

Selon le Gartner, le Big Data (en français méga données ou "Grandes données") regroupe une famille d'outils qui répondent à une triple problématiques : un Volume de données important à traiter, une grande Variété d'informations (en provenance de plusieurs sources, non-structurées, structurées, Open data...), et un certain niveau de Vitesse à atteindre - c'est-à-dire de fréquence de création, collecte, traitement/analyse et partage de ces données [W1].

Le **big data**, ou **méga données** appelées aussi **données massives**, désignent des ensembles de données volumineux qui s'avère difficile à travailler avec des moyens classiques de gestion de base ou de gestion de l'information.

On distingue généralement 2 types de données sous la coupe du Big Data :

- **Les données des entreprises** : les Emails , les documents, les bases de données, tous les historiques de processus métiers (logs), ainsi que tout autre type données structurée, semi-structurée ou non-structurée que l'entreprise produit et stock.

- **Les données en dehors des entreprises** : bases de données externes (publiques ou fournisseurs données), contenus sur les réseaux sociaux ou publiés en ligne, les données (géo localisées ou non) transmises par les objets connectés et les historiques de navigation et de recherche [W2].

✓ Pourquoi le Big Data est si important ?

Le **Big Data** permet, d'abord d'obtenir une représentation assez fidèle de l'interaction des clients avec l'entreprise. En plus offre une meilleure compréhension des intentions des clients à chaque point de contact. Il réduit le risque de perdre ces clients au cours du passage d'un point de contact vers un autre en garantissant ainsi la pertinence de l'information qui leur est donnée [W3].

3. Caractéristiques des Big Data

Big Data couvre 03 dimensions :

3.1 Volume

Le volume décrit la quantité de données produite par des entreprises ou des personnes. Ce qui leur- les entreprises- permet de gérer le volume de données crée quotidiennement en augmentation constante.

3.2 Vitesse.

La vitesse donne la fréquence à laquelle les données sont produits, saisis et partagées. Selon les évolutions technologiques récentes, les consommateurs ainsi que les entreprises produisent plus de données dans des temps plus courts [BM14].

À ce niveau nombreux sont les systèmes d'analyse, de CRM¹, ou autre qui échouent, car les entreprises ne peuvent calculer sur ces données sauf si elles sont données en temps réel.

Ils ne peuvent traiter les données que par vague quelques heures après. Cependant, ces données deviennent obsolètes car le cycle de production de nouvelles données a déjà commencé.

3.3. Variété

La multiplication de données émanant de sources comme les médias sociaux, les interactions *Machine to machine* et les terminaux mobiles, provoque une très grande variété outre les données transactionnelles traditionnelles [BM14].

À ce niveau nombreux sont les systèmes d'analyse, ou autre qui échouent, car les entreprises ne peuvent calculer sur ces données sauf si elles sont données en temps réel.

Ils ne peuvent traiter les données que par vague quelques heures après. Cependant, ces données deviennent obsolètes car le cycle de production de nouvelles données a déjà commencé.

Donc les 3 V (Volume, Vitesse et Variétés) du Big Data sont d'une importance capitale dans une entreprise.

4. Nouveaux enjeux et risques du big data

Le *Big Data* représente de nouveaux enjeux qu'il faut prévoir et en étudier les risques encourus.

De gros volumes en temps réel, sont possibles d'être traités avec le Big Data, permettant par la même de bien contrôler le trafic réseau que d'analyser la qualité de service.

En somme le *Big Data* valorise les données qui, jusque -là, ne pouvaient pas être utilisées telles quelles dans des statistiques sur des périodes prolongées ou même les croisant entre elles.

¹ Customer Relationship Management : c'est l'art d'optimiser les interactions de votre société avec vos clients et vos prospects, Les outils de CRM vous aident à vous concentrer sur le relationnel avec vos clients, vos collaborateurs et vos partenaires.

Négliger le *Big Data* c'est risquer de perdre en compétitivité et d'être en retard sur son marché...Encore s'y précipiter risque aussi de tomber dans l'effet inverse : manque de maîtrise et de fiabilité. Ce qui entraîne un manque de crédibilité pour l'entreprise.

Donc **le risque majeur lié au Big Data** qui apparaît comme un défi technologique est sa mauvaise utilisation. Toujours est-il, il demeure convoité par des politiciens, des scientifiques et des entreprises car il touche plusieurs secteurs d'activités [MC12].

4.1. Enjeux techniques

Les enjeux techniques s'articulent autour de l'intégration, le stockage, l'analyse, l'archivage, l'organisation et la protection des données.

4.2. Enjeux économiques.

« Le Big Data représente 24 milliards de dollars en **2016**, avec une part de stockage estimée à 1/3 de ce montant. » Estimation donnée par le cabinet de conseil dans le marketing **IDC**².

Les spécialistes s'accordent sur le fait que le Big Data sera l'arme économique future pour les entreprises.

Le Big Data permettent :

- D'améliorer les stratégies marketing et commerciale (une démarche d'étude et de réflexion dont le but est de s'approcher au plus près de l'adéquation offre-demande.)
- D'améliorer et entretenir la relation client.
- De fidéliser la clientèle.
- De gagner de nouvelles parts de marché.
- De réduire les coûts logistiques (transport, stocks, informatiques, prestations, frais de personnel, surfaces, équipement,)
- De favoriser la veille concurrentielle (état ou dispositif de surveillance continue des actions et produits des concurrents actuels ou potentiels.)

² IDC pour Indice de Disparité de Consommation est un indice calculé par l' ACFCI (Assemblée des Chambres Françaises de Commerce et d'Industrie). L'IDC permet de corriger une moyenne de consommation nationale d'un produit fournie le plus souvent par l'INSEE pour tenir compte d'une disparité régionale ou locale.

Désormais, avec le Big Data on doit dire « nous avons un client, de quoi a-t-il besoin aujourd'hui ? » au lieu de se demander comme de coutume « nous avons un produit, à qui allons-nous pouvoir le vendre ? »

4.3 Enjeux juridiques.

La protection de la vie privée demeure, le principal enjeu juridique dans le cas où les utilisateurs sont souvent des « produits »[AK,IL13].

5. L'analyse : le point clé du Big data

Les objectifs auxquels répond le Big Data sont :

- **L'extraction** d'informations utiles des données stockées.
- **L'analyse** de ces données.
- **Le restitution** efficace des résultats.
- L'accroissement de **l'interactivité** entre utilisateurs et données.

Remarque :

Google, Facebook, IBM ou JDA Software sont des « entreprises Big Data ».

L'analyse est le point nodal de l'utilisation du **Big Data**. Elle permet de mieux connaître sa clientèle, d'optimiser (améliorer) son marketing, de détecter et prévenir les fraudes, d'analyser son image sur les réseaux sociaux et d'optimiser ses processus métiers [MC12].

6. Processus de chargement et de collecte de données dans big data

La couche responsable du chargement de données dans Big Data, devrait être capable de gérer d'énorme volume de données, avec une haute vitesse, et une grande variété de données. Cette couche devrait avoir la capacité de valider, nettoyer, transformer, réduire (compression), et d'intégrer les données dans la grande pile de données en vue de son traitement. La Figure illustre le processus et les composants qui doivent être présent dans la couche de chargement de données.[NS,HS15][MB15].

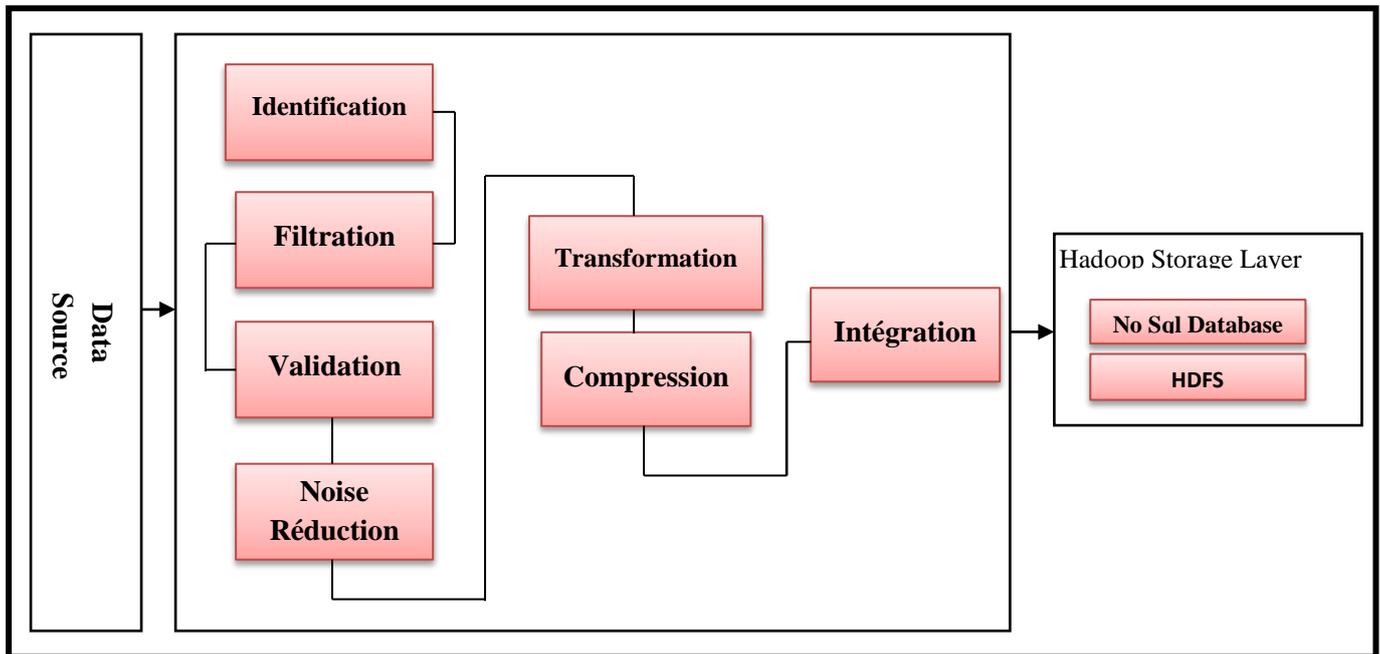


Figure 1.1 : Couche de chargement des données dans le Big Data [NS,HS15].

Big Data collecte les informations pertinentes finales, sans bruit, et les charge dans la couche de stockage (*HDFS* ou *NoSQL base*). Elle inclut dans l'ordre les composants suivants:

- **Identification** : Big Data identifie les différents formats de données connues. En outre, il cible, par défaut, les données non structurées.
- **Filtration et sélection** de l'information entrante pertinente pour l'entreprise.
- **Validation et analyse** des données en permanence.
- **Réduction** de bruit implique le nettoyage des données en supprimant le bruit.
- **La transformation** peut entraîner le découpage, la convergence, la normalisation ou la synthèse des données.
- **Compression** consiste à réduire la taille des données, sans pour autant perdre leur utilité.
- **Intégration** consiste à intégrer l'ensemble des données dans le stockage de données de Big Data (*HDFS* ou *NoSQL base*).

7. technologie des Big Data

7.1. Apache HADOOP

Apache Hadoop (Hight-availability distributed object-oriented platform) est un système distribué qui résout ces problématiques. Premièrement, il propose un système de stockage distribué à travers son fichier HDFS (Hadoop Distributed File System) qui permet de stocker la donnée en la reproduisant. Un cluster Hadoop n'a donc pas besoin d'être configuré avec un système RAID qui devient obsolète. Deuxièmement, Hadoop offre un système d'analyse des données appelée MapReduce qui travaille sur le système de fichiers HDFS afin de faire des traitements sur des volumes de données assez gros [MB14].

Trois principales distributions Hadoop sont aujourd'hui disponibles : Cloudera, Hortonworks, MapR.

7.2. Modèle Mapreduce

Le modèle Map/Reduce est un paradigme de programmation qui permet une scalabilité massive à travers une distribution de données et un traitement parallèle [DJ,GS04]. Il cible la résolution des problèmes liés à la manipulation des données extrêmement volumineuses. Ce modèle a été introduit par Google en 2004 surtout pour construire et gérer son index WEB. Après avoir prouvé son efficacité, il est utilisé par d'autres sociétés telles que Yahoo et Facebook. Son principe de base réside dans le fait de répartir les tâches de traitement (calcul) sur un nombre important de nœuds (machine) selon un modèle de programmation parallèle. Le nombre de noeuds peut atteindre facilement quelques milliers.

Parmi les avantages notables de ce modèle :

- La scalabilité
- La tolérance aux plantages
- La simplicité d'utilisation
- La réduction du coût [AI,AB15].

Un programme MapReduce peut se résumer à deux fonctions Map () et Reduce () .

a. La première, MAP

Va transformer les données d'entrée en une série de couples clef /valeur. Elle va regrouper les données en les associant à des clefs, choisies de telle sorte que les couples clef/valeur aient un sens par rapport au problème à résoudre. Par ailleurs, cette opération doit être parallélisable: on doit

pouvoir découper les données d'entrée en plusieurs fragments, et faire exécuter l'opération MAP à chaque machine du cluster sur un fragment distinct. La fonction Map s'écrit de la manière suivante : $\text{Map}(\text{clé1}, \text{valeur1}) \rightarrow \text{List}(\text{clé2}, \text{valeur2})$.

b. La seconde, REDUCE

Va appliquer un traitement à toutes les valeurs de chacune des clefs distinctes produites par l'opération MAP. Au terme de l'opération REDUCE, on aura un résultat pour chacune des clefs distinctes. Ici, on attribuera à chacune des machines du cluster une des clefs uniques produites par MAP, en lui donnant la liste des valeurs associées à la clef. Chacune des machines effectuera alors l'opération REDUCE pour cette clef. La fonction Reduce s'écrit de la manière suivante : $\text{Reduce}(\text{clé2}, \text{List}(\text{valeur2})) \rightarrow (\text{clé}, \text{valeur})$.

L'exemple classique est celui du WordCount qui permet de compter le nombre d'occurrences d'un mot dans un fichier. En entrée l'algorithme reçoit un fichier texte qui contient les mots suivants
voiture la le elle de elle la se la maison voiture

Dans notre exemple, la clé d'entrée correspond au numéro de ligne dans le fichier et tous les mots sont comptabilisés à l'exception du mot « se ».

Le résultat de la fonction Map est donné ci-dessous.

$(\text{voiture}, 1) / (\text{la}, 1) / (\text{le}, 1) / (\text{elle}, 1) / (\text{de}, 1) / (\text{elle}, 1) / (\text{la}, 1) / (\text{la}, 1) / (\text{maison}, 1) / (\text{voiture}, 1)$

Avant de présenter la fonction Reduce, deux opérations intermédiaires doivent être exécutées pour préparer la valeur de son paramètre d'entrée. La première opération appelée shuffle permet de grouper les valeurs dont la clé est commune. La seconde opération appelée sort permet de trier par clé. A la différence des fonctions Map et Reduce, shuffle et sort sont des fonctions fournies par le Framework Hadoop, donc, il n'a pas à les implémenter [MB14].

Ainsi, après l'exécution des fonctions shuffle et sort le résultat de l'exemple est le suivant :

$(\text{de}, [1]) / (\text{elle}, [1,1]) / (\text{la}, [1, 1,1]) / (\text{le}, [1]) / (\text{maison}, [1]) / (\text{voiture}, [1,1])$

Suite à l'appel de la fonction Reduce, le résultat de l'exemple est le suivant :

$(\text{de}, 1) / (\text{elle}, 2) / (\text{la}, 3) / (\text{le}, 1) / (\text{maison}, 1) / (\text{voiture}, 2)$

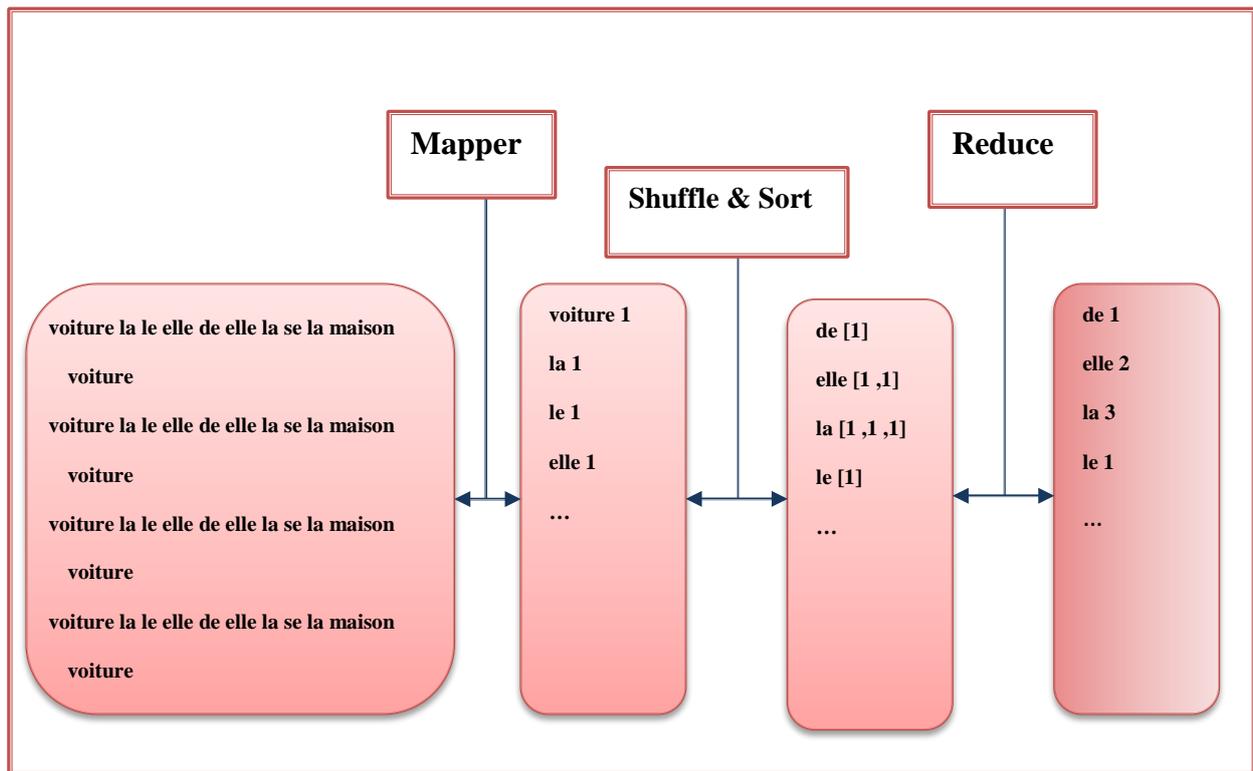


Figure 1.2: Exemple d'un programme MapReduce (WordCount) [MB14]

8. Data Warehouse

Avant de parler de 'Data Warehouse' il nous est apparu très intéressant de le définir; Le « Data Warehouse », en français « Entrepôt de données », forme, dans ces conditions, une structure informatique et une fondation des plus incontournables pour la mise en place d'applications décisionnelles.

Le concept de Data Warehouse, tel que connu actuellement, est apparu pour la première fois en 1980; l'idée consistait alors à réaliser une base de données destinée exclusivement au processus décisionnel.

L'entrepôt de données est destiné à fournir de l'information :

- Thématique. c'est à dire relative à un domaine intéressant le décideur possédant une référence temporelle.
- Sûre. c'est à dire dont la qualité a été vérifiée selon [LHE95].
- Facile d'accès.
- Non volatile, car régulièrement complétée et peu «nettoyée».

9. Les caractéristiques de données d'entrepôts

Un entrepôt de données est une collection de données [LOU10]:

- **Orientée sujet** : Les données organisées par sujet (clients, vendeurs, production, etc..) contiennent purement l'information essentiel à la prise de décision. Ainsi que Les systèmes opérationnels sont plutôt orientés autour des traitements et des fonctions.
- **Intégrée** : Les données, résultantes de différentes sources (systèmes légués) sont généralement structurées et codées d'une manière distinctes. En effet 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.
- **Non volatile** : pour être fixe, en lecture seule, non modifiable pour but de garder la traçabilité des informations et des décisions prises.
- **Archivée et datée** : pour le suivi des évolutions des valeurs des indicateurs à analyser. Les informations stockées au sein du DW ne doivent pas disparaître. Elles deviennent, de ce fait, partie prenante de l'historique de l'entreprise.

Les entrepôts de données sont utilisés dans les systèmes décisionnels. Ils sont le cœur de l'exploitation de l'information de l'entreprise.

10. Composantes d'un entrepôt de données

L'entrepôt de données est composé d'un ensemble de composantes matérielles et logicielles [LOU10]:

10.1. Les applications opérationnelles sources

Ce sont les applications opérationnelles qui capturent les transactions de l'organisation. Les applications sources ne gardent que rarement de données historiées. Un bon entrepôt de données capable de libérer les applications sources d'une bonne partie de leurs responsabilités concernant la représentation du passé.

10.2. La préparation de données

La zone de préparation des données de l'entrepôt est à la fois une zone de stockage et un groupe de processus couramment appelés ETL (Extract/Transform/Load). L'extraction est la première étape du processus d'apport de données à l'entrepôt qui se traduit par la lecture, l'interprétation et la copie des données sources dans la zone de préparation. Ensuite, on passe à la transformation en vue du chargement. Il faut empêcher aux utilisateurs l'accès à la zone de préparation des données. La dernière étape consacre de charger les données, préalablement extraites puis transformées, dans des cibles hétérogènes (le plus souvent des entrepôts de données).

10.3. Les outils d'accès aux données

L'ensemble des outils d'accès aux données constitue le dernier composant majeur d'un environnement d'entrepôt de données. Les outils d'accès aux données constituent l'ensemble des moyens donnés aux utilisateurs pour exploiter la zone de présentation en vue de prendre des décisions basées sur des analyses.

La figure suivante montre ces composantes manière richement détaillée:

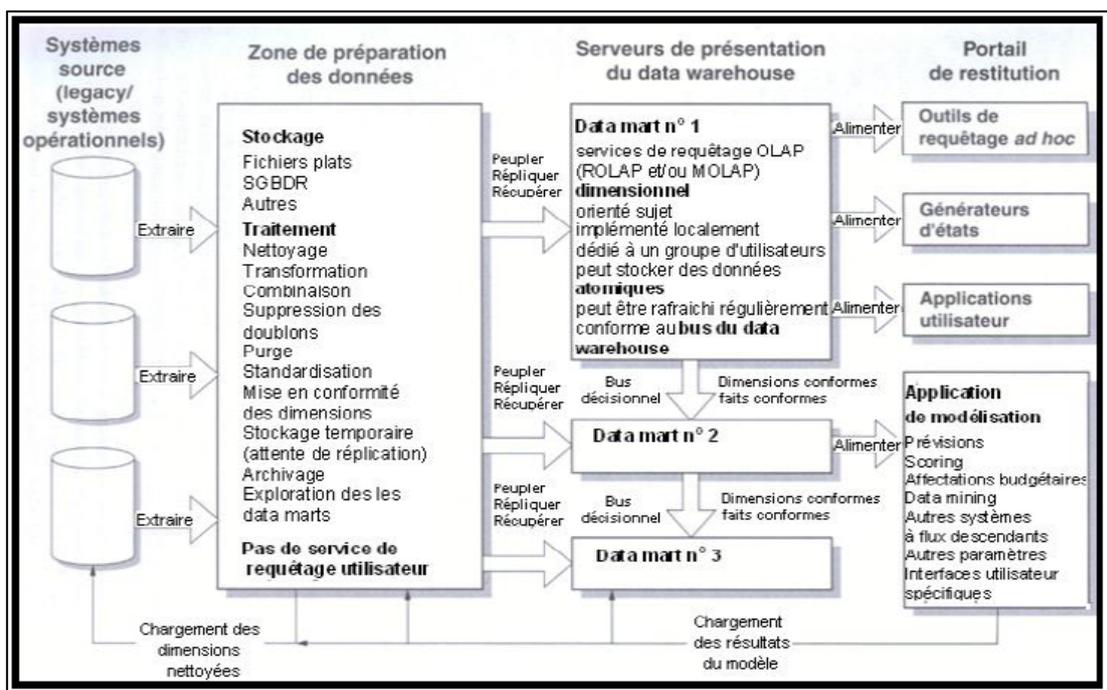


Figure 1.3: Composantes de base détaillées d'un entrepôt de données [LOU10]

11. Les données dans l'entrepôt

La définition des données dans un entrepôt est basée sur des méthodes de vue matérialisée qui sont pris sous plusieurs axes [SEG14]:

- La maintenance incrémentale des vues matérialisées qui se propose de répercuter Immédiatement les mises à jour survenues au niveau des sources de données.
- La configuration de l'entrepôt (sélection des vues à matérialiser)) qui se propose d'amener le groupe de vues à matérialiser dans l'entrepôt de telle manière que le coût de maintenance soit optimal.

12. Principes d'un entrepôt de données

Plusieurs principes, normes et règles régissent un entrepôt de données [LOU10]:

12.1. En amont et en aval (« front-end » et « back-end »)

En amont du DW se place toute la logistique d'alimentation des données de l'entrepôt l'extraction des données opérationnelles, les transformations éventuelles et le chargement dans l'entrepôt : c'est l'ETC ou l'ETL (Extract, Transform Load) ou encore (data maning).

En aval du DW se place tout l'outillage de restitution et d'analyse des données, outils de requêtes, de rapports, de tableau de bord. C'est le BI (Business Intelligence).

12.2. Dimensions et faits

Les notions de dimension et de fait caractérisent le DW. Pour les faits, il s'agit des tables de données dynamiques qui comportent des mesures. Et le dimension c'est les tables de données plus statiques, grossissent beaucoup moins que les tables de fait.

12.3. Architecture d'un entrepôt

Un entrepôt de données est souvent construit selon une architecture en trois couches :

- Un serveur d'entrepôt (serveur de données).
- Un client.
- Un serveur *OLAP* (*On-Line Analytical Processing*).

12.4. Serveurs OLAP (On-Line Analytical Processing)

Les données opérationnelles constituent la source principale d'un système d'information décisionnel. Les systèmes décisionnels complets reposent sur la technologie OLAP, conçue pour répondre aux besoins d'analyse des applications de gestion, Un serveur OLAP (On-Line Analytical Processing) qui peut être de type X_OLAP La différence entre ces technologies s'installe au niveau du système de gestion de la base de données (SGBD) [LOU10]:

- **R_OLAP (Relational OLAP)** : il permet un stockage presque infini de données OLAP ; s'il est relationnel.
- **M_OLAP (Multidimensional OLAP)**: (cube), il y a des limitations quant à la quantité de données M_OLAP ; s'il est multidimensionnel.
- **D_OLAP (Desktop OLAP)**: S'il s'agit de fichiers de stockage sur le poste client, une infime quantité de données est stockée sur le poste (D_OLAP).
- **H_OLAP (Hybrid OLAP)** : S'il supporte et intègre un stockage des données multidimensionnel et relationnel d'une façon semblable afin de profiter des caractéristiques de correspondance et des techniques d'optimisation. [DSB99]

Stratégie	Stockage des données	Traitement des données
MOLAP	Base de données dimensionnelle	Serveur de traitement OLAP
ROLAP	Base de données relationnelle	SQL avancé
DOLAP	Fichier sur le poste client	Client de traitement OLAP
HOLAP	MOLAP pour données sommaires & ROLAP pour données détaillées	

Tableau 1.1 : Alternatives de mise en œuvre OLAP [FAV07]

13. Construction d'un entrepôt de données : [SEG14]

La conception d'un entrepôt de données passe par un cycle de vie qui collectionne les phases suivantes :

13.1. La planification

La visée de cette phase est la préparation du terrain pour le développement de l'entrepôt. Elle garantit les tâches suivantes [SEG14] :

- Déterminer les buts et objectifs de l'entrepôt à développer.
- Evaluer la faisabilité technique et économique de l'entrepôt. cii posant des questions suivant les sources.
- Identifier les futurs utilisateurs de l'entrepôt et leurs rôles.

13.2. Conception et implémentation

Cette phase consiste à développer le schéma de l'entrepôt, et à mettre en place toutes les ressources obligatoires à son implémentation et à son déploiement. Non seulement mais aussi il doit également préparer le déploiement en choisissant une plateforme logicielle, les ressources etc. Cette conception comprend cinq principales étapes [SEG14]:

- **Analyse des besoins** : l'analyse des besoins est une étape essentielle, elle permet la collecte des besoins, analyse des besoins, validation des besoins.
- **La spécification des besoins** : cette étape permet l'ensemble des informations et les données accessibles.
- **Modélisation des données** : permet plusieurs méthodes de modélisation conceptuelle (modéliser le sujet qui est formé de mesure numérique) et modélisation logique (serveur

OLAP), modélisation physique (consiste à implémenter physiquement le schéma de l'entrepôt de données).

14. L'intégration de données

L'intégration de données est un processus permettant de combiner certaines sources afin qu'elles puissent être interrogées à partir d'une interface commune, en éliminant les conflits entre ces données, et les présenter de manière cohérente.

14.1. Les Composants d'un système d'intégration

Un système d'intégration est composé en général en trois couches [BEN15]:

- Une couche de données : contenant l'ensemble des sources de données à intégrer.
- Une couche des adaptateurs ou des chargeurs : qui permettent d'extraire les données et de les représenter dans le schéma global, c'est le moyen avec lequel on peut accéder à une source de données, et avec lequel la source peut interagir avec les autres composants de l'architecture.
- Une couche qui possède le composant qui permet aux utilisateurs d'interroger les différentes sources à partir un schéma global, ce composant peut être un entrepôt de données où toutes les données des sources sont dupliquées.

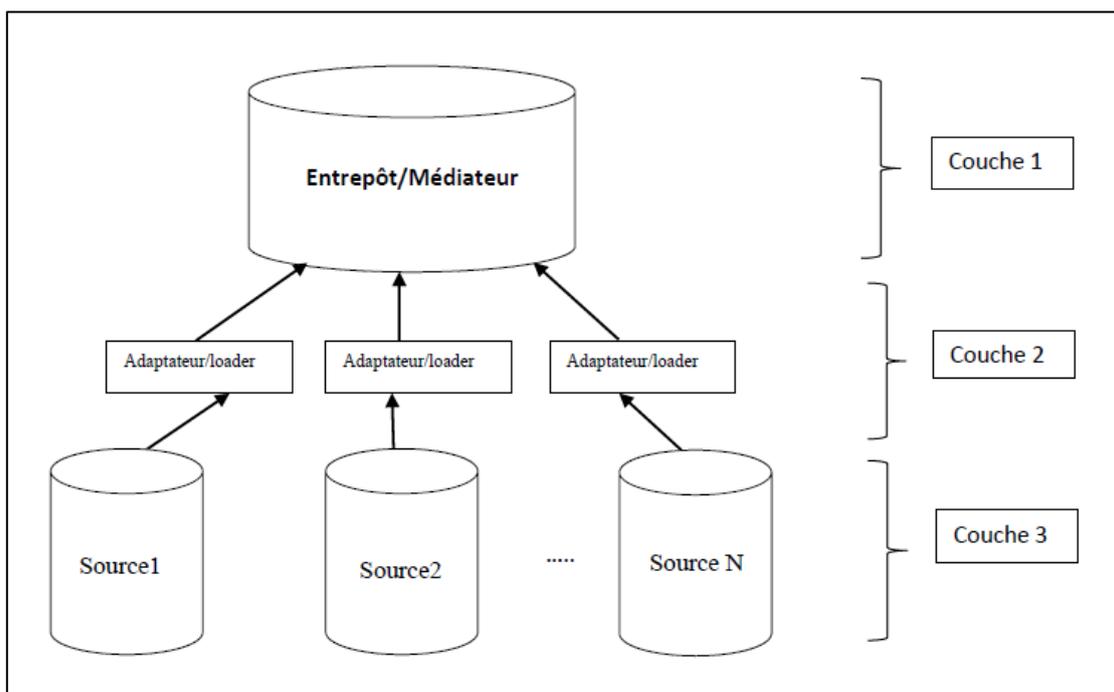


Figure 1.4 : Architecture générale d'un système d'intégration [BEN15].

14.2. Caractéristiques des sources

- **Distribution** : Les données sont stockées sur des supports répartis géographiquement et Offre disponibilité et amélioration des temps d'accès.
- **Hétérogénéité** : L'hétérogénéité concerne les données, les modèles et les langages sémantique et structurelle.
- **Autonomie** : Conception des sources locales avec des modèles de données propres et langage d'interrogation, Interprétation sémantique des données, contraintes, fonctions.
- **Interopérabilité** : Systèmes interopérables permettent l'échange de messages et de requêtes et marchent comme une unité pour une tâche collective, partagent les fonctions et Communiquent aussi avec des composants internes incompatibles.

15. Approches d'intégration de l'entrepôt de données

On trouve trois approches générales pour l'intégration de données [CHA11] :

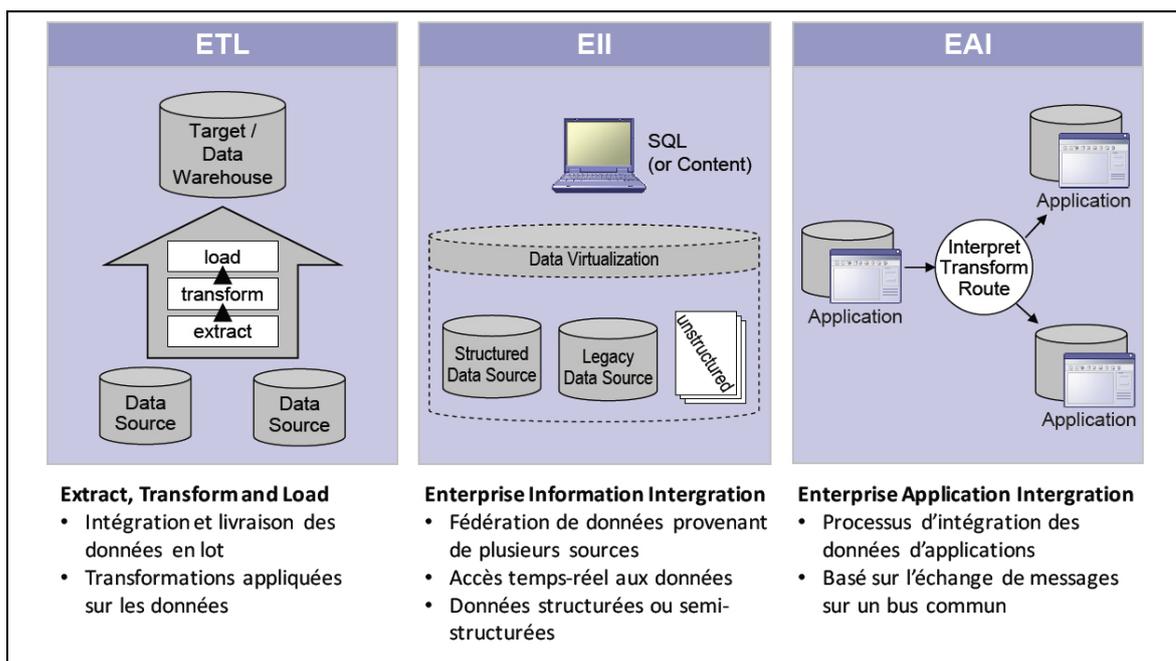


Figure 1.5: Approches d'intégration de l'entrepôt de données [CHA11]

15.1. Enterprise Information Intégration (EII)

a. Caractéristiques:

- Fournit une vue unifiée des données de l'entreprise, où les sources de données installent une fédération.

- Les sources de données dispersées sont consolidées à l'aide d'une BD virtuelle, de façon transparente aux applications utilisant ces données.
- Toute requête à la BD virtuelle est décomposée en sous-requêtes aux sources respectives, dont les réponses sont assemblées en un résultat unifié et renforcé.
- Permet de consolider seulement les données utilisées, au moment où elles sont utilisées (source data pulling).
- Le traitement en-ligne des données peut cependant entraîner des délais importants.

b. Avantages

- Accès relationnel à des sources non-relationnelles;
- Permet de feuilleter les données avec la création du modèle de l'entrepôt de données.
- Accélère le déploiement de la solution.
- Peut être réutilisé par le système ETL dans une itération future.
- Aucun déplacement de données.

c. Inconvénients

- Requier la correspondance des clés d'une source à l'autre.
- Consolidation des données plus complexe que dans l'ETL.
- Surtaxe les système sources.
- Plus limité que l'ETL dans la quantité de données pouvant être traitée.
- Changements limitées sur les données.
- Peut consommer une grande bande passante du réseau.

15.2. Entreprise Application Intégration (EAI)**a. Caractéristiques**

- Approche permettant de fournir à l'entrepôt des données provenant des sources (source data pushing).
- Repose sur l'intégration et le partage des fonctionnalités des applications sources à l'aide d'une architecture SOA.
- Souvent utilisé en temps réel ou en semi-temps réel (NearReal Time).
- L'EAI ne remplace pas le processus ETL, mais permet de simplifier ce dernier.

b. Avantages

- Facilite l'interopérabilité des applications.
- Permet l'accès en (quasi) temps-réel.

- Ne transfère que les données nécessaires.
- Contrôle du flot de l'information.

c. Inconvénients

- Support limité aux transformations et agrégations des données.
- Taille des transactions limitée (en nombre de lignes).
- Développement complexe.
- Gestion complexe de l'intégrité sémantique des données (e.g., règles d'affaires).
- Utilise la bande passante du réseau durant les heures de pointe.

15.3. Extract, Transform and Load (ETL)**a. Caractéristiques**

Permet la consolidation des données à l'aide des trois opérations suivantes:

- Extraction: identifier et extraire les données de sources ayant subi un changement depuis la dernière exécution.
- Transformation: appliquer plusieurs transformations aux données pour les nettoyer, les intégrer et les agréger.
- Chargement: insérer les données modifiées dans l'entrepôt et gérer les changements aux données existantes (ex: stratégies SCD).
- Traite normalement de grandes quantités de données en lots cédulés.
- Est surtout utilisé avec les entrepôts de données et les comptoirs de données.

b. Avantages

- Optimisé pour la structure de l'entrepôt de données.
- Peut traiter de grandes quantités de données dans une même exécution (traitement en lot).
- Permet des transformations complexes et agrégations sur les données.
- La cédule d'exécution peut être contrôlée par l'administrateur.
- La disponibilité d'outils GUI sur le marché permet de progresser la productivité.
- Permet la réutilisation des processus et transformations.

c. Inconvénients

- Processus de développement long et coûteux.
- Gestion des transformations obligatoires.

- Exige de l'espace disque afin effectuer les changements.
- Exécuté séparément du besoin réel.
- Latence des données entre la source et l'entrepôt.
- Unidirectionnel (des sources vers l'entrepôt de données).

	ETL	EII	EAI
Flot de données	Unidirectionnel (sources à l'entrepôt)	Bidirectionnel	Bidirectionnel
Mouvement De données	Lots cédulés	Au moment de la requête	Déclenché par la transaction
Latence	Journalier à mensuel	Temps-réel	Quasi temps-réel
Transformations/ agrégations des données	Grande capacité	Moyenne capacité	Faible capacité
Volume Des données	Grand (millions Ou milliards de lignes)	Moyen (10,000–1,000,000 de lignes)	Petit (100-1000 lignes)

Tableau 1.2 : Comparaison entre les approches d'intégration [CHA11]

Conclusion

Le concept « Big Data » est apparu comme une réponse à des besoins grandissants dans le domaine décisionnel. Son adaptabilité et sa capacité de fournir les données nécessaires à une bonne analyse, ont fait de lui un atout majeur et incontournable pour toute entreprise soucieuse du suivi de ces performances.



Chapitre 03:

Contribution : modèle d'intégration avec ETL et service web

1. Introduction
2. Les usages d'entrepôt de données dans la Médecine
 - 2.1. L'intégrer des sources des données patientes hétérogènes
3. Intégration des données de prescription dans un entrepot de données
4. Entreposage de données complexes pour la médecine personnalisée
5. Notre proposition
 - 5.1. Sources Hétérogènes
 - 5.2. Service Web
 - 5.3. Data WareHouse
 - 5.4. OLAP
6. L'implémentation de modèle
 - 6.1. Technologies XML
 - 6.2. Sources hétérogènes
 - 6.3. Service web et ETL

Conclusion

1. Introduction

L'informatique occupe une place de choix dans le domaine médical, au point de vue technicité et prise en charge du patient. Les systèmes informatiques médicaux sont désormais omniprésents, visibles dans le cas de systèmes de dossiers médicaux et s'efface dans le cas des systèmes embarqués qui contrôlent les dispositifs techniques.

Beaucoup de questions importantes sont posées D'abord, à propos du stockage et de la réutilisation de l'information biomédicale, épidémiologique et médicale. Ensuite sur la modélisation des savoirs et savoir-faire médicaux. Encore sur la gérance de toutes les spécialités médicales. Aussi sur le partage de l'information médicale sans pour autant dévoiler le secret médical. Enfin, sur le passage du signe vers le diagnostic. En définitive sur l'évolution des connaissances des systèmes d'information.

Ainsi, ce chapitre sera consacré à la description De l'informatique médicale et leur Importance dans le secteur de la santé.

2. L'informatique médicale

2.1. Définition

Les systèmes informatiques serviraient plus à veiller au respect de la réglementation, à protéger le personnel médical des risques de procès, et à épargner du temps et de l'argent qu'à améliorer l'offre de soin, la qualité des diagnostics et la coordination sanitaire.

L'informatique de santé fournit rarement une vue d'ensemble des données relatives aux patients, alors que les professionnels de santé documentent les soins prodigués en vue de respecter la réglementation, et d'éviter tout risque de procès plutôt que pour améliorer la qualité des soins[JMM09].

L'informatique médicale est la science et l'art de la modélisation et de l'enregistrement des concepts et des événements cliniques du monde réel dans des données calculables utilisées pour dériver des informations utiles, basées sur l'expertise en médecine, en science de l'information, en technologie de l'information et dans l'étude savante de Les problèmes qui ont un impact sur l'utilisation productive des systèmes d'information par le personnel clinique.

L'objectif de l'informatique médicale (**IM**) est la contribution à l'amélioration de la qualité des soins grâce à une meilleure maîtrise (accès, mémorisation, organisation, partage) de l'informatique médicale liée notamment:

- À la « forme » sous laquelle elle est transmise : texte, valeurs numériques, signaux physiologiques, images...
- À la valeur qu'elle peut prendre : variable quantitative (pouls, tension artérielle...) ou qualitative (réflexes présents ou diminués...) - aux méthodes et appareils de mesure (variabilité analytique)
- Aux observateurs (variabilité intra- et inter-observateurs)
- Au sujet observé (variabilité intra- et interindividuelle)
- À la nature de l'information (données cliniques et/ou d'imagerie, données des laboratoires, bases de connaissances, ouvrages médicaux, revues...) [DP, FM05].

2.2. Les grands domaines de l'informatique médicale

Les progrès médicaux concernant les traitements pharmacologiques ont aidé à soigner plus de personnes. Les preuves issues d'essais cliniques ont permis d'établir des recommandations. Cette médecine basée sur les preuves sert de pierre angulaire à la bonne pratique médicale. Or, un grand nombre de patient ne répond pas aux traitements proposés, et cela a pour cause les caractéristiques de chacun dans la capacité réactive avec les médicaments.

Pour mieux cibler l'approche diagnostique des patients, il faudra mieux les caractériser dans une médecine dite personnalisée. Cette caractérisation fera appel aux progrès de la pharmacogénétique. La médecine factuelle et la médecine personnalisée devront être complémentaires bien qu'elles entraînent d'inévitables défis éthiques face auxquels le praticien, les autorités et l'industrie pharmaceutique devront rester vigilants.

2.2.1. L'EBM (Evidence Based Medicine)

La médecine factuelle, appelée **EBM** par les auteurs anglo-saxons est basée sur les preuves [SAJ15]. Ces preuves sont issues des résultats d'études d'intervention sur de larges cohortes dont la méthodologie est éprouvée et reconnue. L'EBM est née de la volonté de certaines équipes d'enseigner la médecine selon un plan plus systématique. Elle est ensuite devenue une aide au jugement clinique dans la pratique médicale.

L'EBM est définie comme l'utilisation systématique des données de la science dans les décisions visant les soins aux malades. La médecine factuelle se base sur les données issues des essais cliniques d'intervention permettant de respecter la méthodologie et les aspects éthiques, mais ces essais ont des limites [RW,DA95].

2.2.2. La médecine personnalisée

Cette médecine a été développée depuis une vingtaine d'années par la société suisse Roche, l'un des leaders internationaux de l'industrie pharmaceutique commercialisant l'Herceptine¹, un traitement anti-cancéreux, développé pour la prise en charge de certaines formes du cancer du sein, qui a été à la base des thérapies ciblées [I.SM05].

¹ est un médicament issu des biotechnologies, Il s'agit tout simplement d'un anticorps qui s'attaque directement aux cellules tumorales et plus particulièrement à certaines tumeurs qui produisent une protéine, appelée HER2, en trop grande quantité..

. 2.3. Importance de l'informatique dans le secteur de la santé

L'informatique devient un moyen d'aide à la décision: elle présente, trie et contextualise ces données. De plus, la connaissance scientifique explose, seule l'informatique peut la gérer. Enfin, la santé n'a pas de prix mais elle a un coût.

L'informatique permet un retour et un archivage d'informations. C'est précieux pour les instances qui rémunèrent le personnel soignant [W4].

2.3.1 La disponibilité de l'information

Internet offre des informations sur tout. De nouvelles découvertes dans le domaine médical sont disponibles sur l'Internet et facilement commander des médicaments par Internet. Envoyer des rapports médicaux par e-mail et obtenir des suggestions par un médecin se trouvant dans l'autre bout du monde [W5].

2.3.2 Gestion de base de données

Nous vivons dans un monde de plus en plus connecté, aux outils informatiques qui deviennent de plus en plus obligatoires dans certains domaines tels que l'enseignement et la médecine.

Avant de parler de la gestion de base de données dans les cliniques, il faudra définir la notion "base de donnée ",qui permet de stocker des données d'une manière organisée pour être utilisées d'une façon structurée.

Ainsi, cette notion est couplée à celle de **réseau**, pour mettre en commun ces informations, d'où le nom de base. On parle généralement de système d'information pour désigner ce qui regroupe les outils mis en place servant à partager des données. Afin de gérer les bases de données, il a un ensemble de services (applications logicielles) (**le SGBD**) pour:

- Permettre l'accès des données de façon simple.
- Autoriser un accès aux informations à plusieurs usagers.
- Manipuler les données présentes dans la base de données (insérer, supprimer, modifier) par plusieurs usagers simultanément.

La gestion de base de données dans les cliniques et les hôpitaux devient facile avec l'arrivée de l'ordinateurs qui aide à garder la trace de tous les patients antécédents.

Ainsi, toutes les informations sont disponibles au clic d'un bouton pas seulement mais aussi les enregistrements de rapports médicaux de millions de patients.

En outre, le stockage des informations est un moyen écologique, divers sites sont disponibles en quelques minutes sur Internet [W5].

2.3.3 Technologie avancée

La médecine et les technologies avancées sont étroitement liées d'où la conception de logiciels et machines automatiques dans le domaine médical permettant le traitement thérapeutique. Les outils de diagnostic exigent la présence d'ordinateurs [SB12].

« La médecine, la pensée du médecin et du patient sont envahies par la technologie, mais we can't do without it. On ne peut pas ! Mais ce n'est pas facile ça » .

2.3.4 Imagerie médicale

Aujourd'hui, les techniques d'imagerie sont nombreuses et complémentaires. Elles ont évolué à partir de grandes découvertes physiques du 20^{ème} siècle, les rayons X et les ondes radio, la radioactivité naturelle et artificielle qui s'appuient sur les progrès de la médecine et les avancées scientifiques et informatiques [AG17].

2.4. Principales applications de l'informatique médicale

L'informatique est utilisée, dans le domaine médical, pour permettre:

- L'informatisation du dossier patient.
- L'informatisation des laboratoires
- Les systèmes d'enseignement : l'enseignement assisté par ordinateur (EAO) ¹
- Pubmed : portail WEB sur internet pour l'interrogation des banques de données bibliographiques (medline) de la bibliothèque nationale américaine (National Library of Medicine).
- Le Programme de Médicalisation des systèmes d'information [LF12].

¹ EAO :est une spécialité informatique qui regroupe les logiciels permettant l'aide à l'apprentissage dans des domaines divers, ainsi que les outils utilisés pour créer ces programmes.

2.4.1. Systèmes documentaires et banques d'information

Les systèmes documentaires informatisés sont de deux types :

- Certains sont conçus comme des index d'information contenant une partie de l'information: **appelés bases de données de recherche bibliographique.**
 - ✓ Medline est la plus connue. Elle permet d'effectuer une recherche sur internet en version gratuite ¹, ou payante ², cette dernière offre des services supplémentaires.
 - ✓ Les maisons d'édition (Elsevier, Masson, Springer...) mettent en ligne les journaux avec un accès payant.
 - ✓ Les bases de données de recherche bibliographique dites analytiques permettant l'accès aux résumés des articles est gratuit.
- D'autres visent à fournir l'information recherchée: **ce sont les bases de connaissance.**

L'Université Médicale Virtuelle Francophone (UMVF) en est l'exemple. Mais il existe sur le net plus de 20000 sites médicaux de qualité.

- Un certain nombre de portails médicaux et de moteurs de recherche spécialisés :
 - ✓ Portails médicaux : (CHU de ROUEN) ³.
 - ✓ Moteurs de recherches médicaux : Sumsearch, Tripdatabase... [DP, FM05].

2.4.2. Le dossier médical informatisé (DMI)

Le DMI améliore le stockage, le regroupement des données et la communication entre les intervenants car il constitue un Outil de mémorisation dans la recherche médicale si les informations pathologiques concernant le même malade sont disparates: une partie sera détenu par l'hôpital, l'autre par le médecin et l'autre par la compagnie d'assurance.

Des procédures automatiques de connexion à des bases de données sont plus accessibles que les dossiers classiques, les DMI sont mieux protégés et permettant d'assurer la confidentialité des données. L'informatisation et la communication plus faciles au niveau des cabinets médicaux qu'à l'hôpital.

¹ PubMed, NLM Gateway, BioMedNet, Medscape.

² Ovid Technologies et SilverPlatter.

³ Centre Hospitalier Universitaire: est un établissement hospitalier dans lequel sont prodigués des soins par des professionnels de santé confirmés et des étudiants.

La carte mémoire utilisée en Europe permet la mémorisation d'informations critiques et des informations d'identification servant au paiement des actes médicaux. Son utilisation ainsi que la disponibilité de lecteurs dans le réseau médical posent problème [DP, FM05].

2.4.3. Systèmes d'aide à la décision médicale

Les systèmes d'information hospitaliers, les systèmes de réseaux de soins, les dossiers médicaux informatisés, les bases de connaissance facilitent l'accès aux données du patient, et permettent d'en faire une présentation synthétique pour ensuite les confronter aux informations figurant sur les bases de connaissances.

Les systèmes informatiques d'aide à la recherche ont évolué dans le domaine de la biologie moléculaire et de l'étude des gènes à partir de la gestion de grandes bases de données pouvant mémoriser plusieurs milliers de fragments du génome humain [DP, FM05].

Le système d'aide à la décision médicale est un ensemble organisé d'informations, crée pour aider un praticien dans son raisonnement afin de diagnostiquer et de mettre en place la thérapeutique adéquate, en réalisant un dialogue entre l'homme et la machine.

Les (SADM) sont des applications informatiques fournissant aux cliniciens au moment opportun les informations décrivant la situation d'un patient ainsi que les connaissances relatives à cette situation, afin d'améliorer la thérapeutique de ce dernier [AB15].

3. Médecine personnalisée

3.1. Définitions

La médecine personnalisée, ou médecine de précision, vise à traiter chaque personne de façon individualisée en fonction de ses spécificités génétiques et aux caractères particuliers de sa maladie[W6].

Cette médecine, admise et reprise par l'EMA (l'Agence Européenne du Médicament) [EMA11] consiste à :

« Donner au bon patient le bon traitement, chaque médicament étant donné à la bonne dose au bon moment », ce à quoi on peut ajouter « et pour la bonne durée ».

Selon la Food and Drug Administration (FDA) [FDA13]

« Il s'agit d'ajuster le traitement médical aux caractéristiques individuelles, besoins et préférences de l'individu à toutes les étapes du soin (care), incluant la prévention, le diagnostic, le traitement et le suivi ».

Contrairement à une médecine qui serait de taille unique, la médecine personnalisée est souvent présentée comme une médecine « sur-mesure ». Tous les niveaux de l'acte médical y sont concernés : Le diagnostic moléculaire, l'aide des biomarqueurs (Une caractéristique biologique mesurable liée à un processus normal ou non.) et les modalités thérapeutiques détaillées

a. Les biomarqueurs

Les biomarqueurs sont des molécules ou des protéines surexprimées ou anormalement absentes dans certains types de tumeurs. Ces marqueurs peuvent être annonciateurs (prédictifs) d'une réponse positive à une thérapie ciblée ou encore de l'évolution d'une tumeur voire de sa sévérité.

Rôle des biomarqueurs :

- ✓ Établir un diagnostic précis.
- ✓ Choisir un traitement ciblé quand il est disponible.
- ✓ Établir un pronostic sur l'évolution d'une tumeur.
- ✓ Suivre l'efficacité d'un traitement.

- ✓ Suivre la résistance d'une chimiothérapie.
- ✓ Évaluer les risques d'apparition de métastases.

Plusieurs sont déjà connus, comme les récepteurs hormonaux ou encore le récepteur **HER2**¹ (Le récepteur **HER2** fait partie de la famille des récepteurs EGFR (Epithelial Growth Factor Receptor), est impliqué dans la régulation de la prolifération et de la survie des cellules et, en particulier, des cellules cancéreuses). **HER2** permet de prédire respectivement la réponse à un traitement hormonal et à la thérapie ciblée Herceptin®² (indiqué dans le traitement de patients adultes atteints d'un cancer du sein...).

b. La médecine stratifiée

Consiste à définir des sous-populations (groupes ou proportions de patients) en fonction du « sous-type » de maladie diagnostiquée chez un patient. Exemple : certains cancers du sein sont récepteurs hormonaux –positifs », d'autres sont « HER 2 –positif », d'autres encore ne sont ni l'un ni l'autre.

Le cancer du sein peut être lié aux œstrogènes ou à la progestérone. Les cellules de cancer du sein qui sont ER+³ et/ou PR+ sont susceptibles de répondre aux médicaments qui bloquent l'action des œstrogènes ou de la progestérone. Presque 60% des cas de cancer du sein répondent à ces médicaments. Les mêmes médicaments ne pourraient être efficaces que dans environ 5 à 10% des cas si la tumeur n'est pas ER+ ou PR+⁴

La médecine de précision est basée sur une information moléculaire qui améliore la précision du diagnostic, et donc la façon dont les maladies sont traitées. Il ne s'agit donc, une fois encore, que de stratification de la population pour prescrire à bon savoir le médicament et en ce sens, médecine de

¹ Le récepteur HER2 fait partie de la famille des récepteurs EGFR (pour Epithelial Growth Factor Receptor). L'Epithelial Growth Factor est impliqué dans la régulation de la prolifération et de la survie des cellules et, en particulier, des cellules cancéreuses.

² Herceptin est un médicament qui contient le principe actif trastuzumab. Il est disponible sous la forme d'une poudre destinée à être reconstituée en solution pour perfusion (goutte-à-goutte) dans une veine et sous la forme d'une solution pour injection sous-cutanée. Herceptin est utilisé pour traiter les types de cancer

³ Premier Groupe de tumeurs possédant des récepteurs aux œstrogènes, Elles concernent donc surtout les cellules luminales (lobulaires ou canalaire), celles qui vont fabriquer le lait . Ces cellules cancéreuses possèdent d'ailleurs également des récepteurs à la progestérone et à la prolactine, et se caractérisent par des gènes qui gèrent la fabrication des protéines caractéristiques des cellules luminales

⁴ la progestérone cellules cancéreuses ont des récepteurs PR+, Quand les cellules cancéreuses ont ces récepteurs, les hormones peuvent s'y fixer et aider ces cellules à croître.

précision peut être considérée comme synonyme de médecine stratifiée, et non de médecine personnalisée qui repose sur des modalités très diverses, allant au-delà de la seule stratification de la population [KNA13].

4. Big Data et santé

Comme tous les domaines, les soins de santé et l'industrie médicale dans son ensemble sont peu à peu bouleversés par le Big Data. Grâce à la collecte et l'analyse de données massives, les techniques de prévention, de traitement, de diagnostic et de suivi des patients évoluent à vive allure.

Pendant de longues années, les soins de santé, la recherche et les découvertes médicales ont reposé sur la collecte et l'analyse de données. Les professionnels de la santé cherchaient alors à comprendre qui tombe malade, comment, et pourquoi. Aujourd'hui, grâce aux nombreux capteurs des smartphones et l'augmentation de la quantité et de la qualité de données, les possibilités de découvertes révolutionnaires sont en forte croissance.

4.1. Big Data et médecine personnalisée

Le Big Data assure également la jonction entre de vastes quantités de données avec les soins de **patients individuels**. Des algorithmes comme IBM **Watson** ont prouvé leur capacité à analyser de larges quantités de textes. À présent, **IBM** cherche à produire une interface qui permettrait à **Watson** d'analyser les recherches médicales existantes sur n'importe quel sujet afin de synthétiser et de résumer ces informations pour les médecins.

Par la suite, il est fort probable que les médicaments et traitements prescrits par les médecins soient également établis avec l'aide du **Big Data**. Les données permettent aux chercheurs de choisir les meilleurs sujets pour des essais cliniques, et les algorithmes leur permettent d'analyser les résultats. Récemment, des accords de partage de données entre les géants de l'industrie pharmaceutique ont conduit à de grandes découvertes. **Par exemple**, le fait que la Desipramine, initialement utilisée comme antidépresseur peut également servir à soigner certains cancers du poumon.

Dans un futur proche, les docteurs pourraient également choisir d'utiliser **la médecine personnalisée** en guise de traitement. Ceci implique de créer des médicaments sur mesure en fonction du code génétique d'un patient. Pour ce faire, l'empreinte génétique, le mode de vie et l'environnement d'un patient sont comparés à des milliers d'autres pour prédire la maladie et déterminer le traitement le

plus adapté. En outre, le **Big Data** est également utilisé pour traquer, et **analyser et soigner les épidémies mondiales comme Ebola ou Zika [BL16]**.

En conséquence, le médecin sera capable de choisir les meilleurs traitements pour un patient individuel en se basant sur les larges quantités de données disponibles sans avoir besoin d'effectuer de longues recherches.

Toutes ces technologies ne représentent qu'un début. Au fur et à mesure, **de nouveaux traitements et diagnostics améliorés sauveront plus de vies et permettront de soigner plus de maladies**. À n'en point douter, le futur de la médecine et des soins de santé repose sur **le Big Data** et l'intelligence artificielle [BL16].

5. Système d'aide à la décision médicale

5.1. Décision médicale

En médecine, la décision constitue le centre de l'acte médical. Elle consiste à diagnostiquer, proposer un traitement ou le différer. En conséquence, plusieurs applications d'aide à la décision ont été développées dans ce domaine pour soutenir le personnel de santé et d'aider à utiliser les moyens d'aide à la décision [DP, FM05].

5.2. Définition d'un système d'aide à la décision médicale

Les systèmes d'aide à la décision médicale (SADM) sont des outils informatiques servant à fournir aux cliniciens les informations propres à la situation du malade et les connaissances filtrées pour améliorer la qualité des soins [BES09].

Il existe des (SADM) pour l'ensemble des activités médicales et la majorité des spécialités médicales. Ces systèmes proposent des services pour les différentes catégories de médecins et les différentes modes d'exercices. En outre ces SADM ont été développés à destination des patients afin qu'ils soient mieux informés sur leur maladie et les soins qui pourraient leur être proposés en vue de décision partagée [BS,JB15].

5.3. Différents types de systèmes d'aide à la décision médicale

Selon **Shortliffe**: "Les SADM désignent de nombreuses solutions partageant la même visée d'amélioration des pratiques. Or chaque approche possède ses propres caractéristiques"[SEH 87].

Les SADM sont répartis en 3 catégories principales :

5.3.1. les systèmes d'aide indirecte à la prise de décisions ou d'assistance documentaire

Malgré le rôle réduit de ces systèmes à la simple mise à disposition des données, ces SADM de gestion de l'information médicale avancent les données patient et les sources bibliographiques pour faciliter au médecin l'accès aux informations pour la prise en charge du malade [AG15].

5.3.2. les systèmes d'alerte ou de rappels automatiques

Certains systèmes permettent au médecin de ne pas commettre des erreurs ou de prendre en compte des éléments importants pour la décision car ils sont actifs et directement impliqués dans la décision médicale. L'assistance n'est pas une aide au raisonnement du cas patient, mais plutôt un aide fournissant une information pertinente dans une situation clinique.

Ces systèmes, comme les précédents, ne raisonnent pas véritablement.

Par exemple, le rappel des valeurs normales des résultats d'examens de biologie et l'utilisation d'une typographie permettant d'attirer l'attention sur les valeurs anormales ou encore l'émission d'un message de mise en garde devant une association de médicaments déconseillée constituent une aide simple dont l'utilité est primordiale [AB15].

5.3.3 les systèmes consultants ou "Aide centrée patient" (Patient-centered support)

Ces systèmes s'appuient sur les données patient et sur une base de connaissance pour fournir des conseils adaptés à la prise en charge de ce patient. Les recommandations "centrées-patient" peuvent guider le médecin vers un diagnostic (on parle alors de SADM orienté diagnostic) ou vers la meilleure prise en charge (SADM orienté thérapeutique) [AG15].

6. Les systèmes experts pour l'aide à la décision médicale

Au cours des années 70, avec l'avènement de l'intelligence artificielle dont l'objectif était la recherche de moyens susceptibles de doter les systèmes informatiques de capacités intellectuelles

comparables à celles des humains, de nouveaux systèmes d'aide à la décision médicale, appelés systèmes experts [BS,JB15].

6.1. Définition

Les systèmes experts tentent de simuler le savoir-faire, la façon de raisonner des experts dans un domaine donné bien délimité et précis et la mise des connaissances apprises par le processus à la disposition des utilisateurs ou experts du domaine. Ils doivent réaliser à la fois un traitement d'information et un raisonnement, ce qui signifie en termes médicaux, une stratégie de diagnostique et/ou thérapeutique [DAR03].

Donc le rôle d'un système expert médical est de fournir une aide médicale sous une forme appropriée à partir de symptômes préalablement établis.

Les systèmes experts sont des systèmes à base de connaissances qui visent à simuler le raisonnement des experts engagés dans une démarche décisionnelle. Les connaissances nécessaires à la résolution d'un problème sont ainsi dissociées du raisonnement à mettre en œuvre pour produire la solution. Classiquement un système expert est constitué de 3 modules principaux :

- ✓ la base de connaissances contenant les connaissances médicales théoriques et empiriques servant à diagnostiquer ou à soigner dans les domaine d'expertise.
- ✓ la base de faits caractérisant le patient dont on cherche le diagnostic ou la prise en charge thérapeutique,
- ✓ le moteur d'inférences articulant les connaissances de la base de connaissances pour raisonner sur la résolution du problème posé dans la base de faits [BS,JB15].

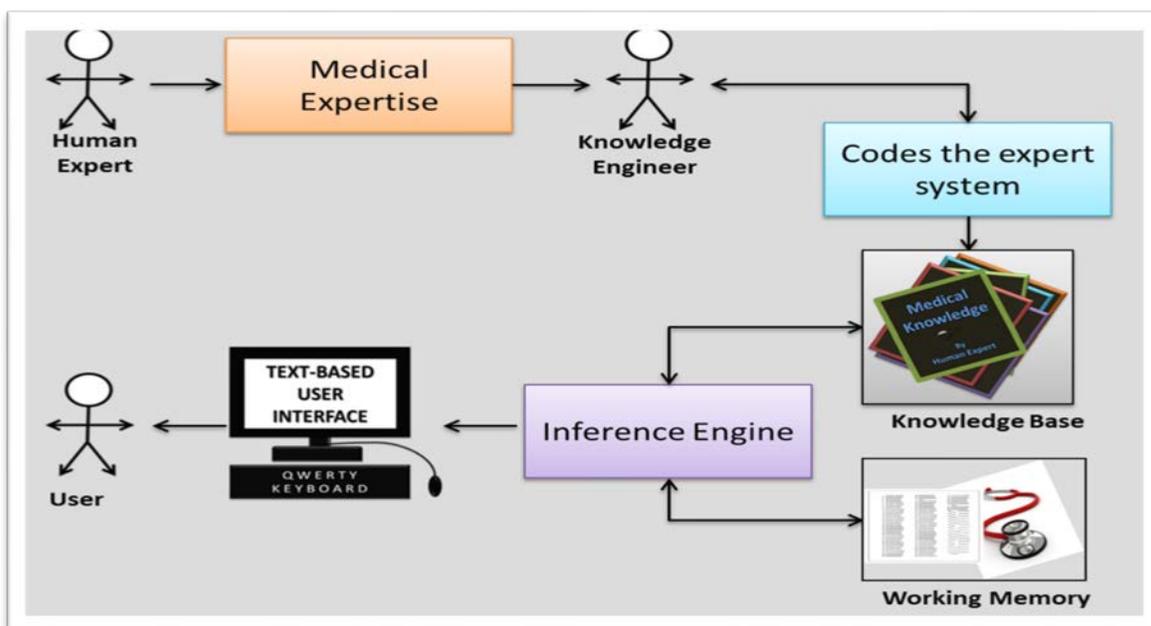
6.2. Architecture d'un Système Expert

Le système expert représente les connaissances d'expertise en tant que règles et faits au sein de l'ordinateur. Ces règles et faits peuvent être utilisés lorsque nécessaire pour résoudre des problèmes. Cependant, un système basé sur la connaissance saisit les connaissances d'un expert humain et utilise ces connaissances pour résoudre un problème réel en temps réel.

6.2.1. composants du système expert

Le système expert se compose de quatre éléments principaux: base de connaissances, mémoire de travail, moteur d'inférence et interface utilisateur.

- ✓ **Base de connaissances:** Le cœur d'un système expert basé sur la connaissance. Il contient une collection de connaissances sur le domaine du problème. Il stocke toutes les informations pertinentes utilisées par le système expert.
- ✓ **Mémoire de travail:** Semblable à un système de base de données relationnelle. Il contient toutes les connaissances sur le problème fourni par l'utilisateur final.
- ✓ **Moteur d'inférence:** Le cerveau du système expert basé sur la connaissance. C'est une analogie avec le raisonnement humain. Il met en œuvre le processus de raisonnement de l'intelligence artificielle. Son rôle est de travailler avec les données disponibles du système et de l'utilisateur pour trouver une solution au problème.
- ✓ **Interface utilisateur basée sur le texte:** L'intermédiaire qui permet la communication entre l'utilisateur et le système. Le but de l'interface utilisateur est de faciliter l'utilisation du système expert par les développeurs, les utilisateurs et les administrateurs [KB, KM12].



Architecture de système expert [KB, KM12].

Conclusion

Compte tenu des limites de L'outil informatique, un travail en parallèle prenant en considération les aspects émotionnels et inconscients de la pratique du soin. Ce qui augure des résultats probants si le jumelage de la technicité performante et l'attention du patient s'associeront.



Chapitre 02:

Informatique médicale et Médecine Personnalisée

1. Introduction
2. L'informatique médical
 - 2.1. Définition
 - 2.2. Les grands domaines de l'informatique médicale
 - 2.2.1. L'EBM(Evidence Based Medicine)
 - 2.2.2. La médecine personnalisée
 - 2.3. Importance de l'informatique dans le secteur de santé
 - 2.3.1. La disponibilité de l'information
 - 2.3.2. Gestion de base de données
 - 2.3.3. Technologie avancée
 - 2.3.4. Imagerie médicale
 - 2.4. Principales applications de l'informatique médicale
 - 2.4.1. Systèmes documentaires et banques d'information
 - 2.4.2. Le dossier médical informatisé (DMI)
 - 2.4.3. Systèmes d'aide à la décision médicale
3. Médecine personnalisée
 - 3.1. Définitions
4. Big Data et santé
 - 4.1. Big Data et médecine personnalisée
5. Système d'aide à la décision médicale
 - 5.1. Décision médicale
 - 5.2. Définition d'un système d'aide à la décision médicale
 - 5.3. Différents types de systèmes d'aide à la décision médicale
 - 5.3.1. les systèmes d'aide indirecte à la prise de décisions ou d'assistance documentaire
 - 5.3.2. les systèmes d'alerte ou de rappels automatiques
 - 5.3.3. les systèmes consultants ou "Aide centrée patient" (Patient-centered support)
6. Les systèmes experts pour l'aide à la décision médicale
 - 6.1. Définition
 - 6.2. Architecture d'un Système Expert
 - 6.2.1. composants du système expert

Conclusion

1. Introduction

Dans ce chapitre, nous présentons notre modèle d'intégration des données, avec les processus ETL avec les services web, En fonction du travail connexe, nous avons essayé de résoudre l'hétérogénéité en utilisant les services Web et XML, nous avons utilisé Microsoft Visual Studio comme environnement de développement.

2. Les usages d'entrepôt de données dans la Médecine

La médecine moderne est devenue presque inconcevable sans l'utilisation des données personnelles numérisées. L'émergence de l'e-santé, la télémédecine, les NBIC (nanotechnologies, biotechnologies, informatique, et sciences cognitives), et des Big Data modifient la prestation de santé, la relation médecin patient, et la compréhension scientifique du corps humain et des maladies [MAR14]:

- **Pour l'individu:** Du point de vue médecin, il y a des domaines très différents, L'aide au diagnostic, La médecine ciblée et personnalisée.
- **Les populations :** Les professionnels de la santé, avec la possibilité d'un diagnostic plus rapide et précis, voire d'un traitement totalement personnalisé pour les patients et Les collectivités afin de remplir leur rôle de vigilances et de veille sanitaire.
- **La recherche :** Détection de cohortes et les études de faisabilité clinique.
- **Médico-administratif :** organisation raisonnée des soins, pilotage des activités, analyse des trajectoires de santé.
- **Pour l'enseignement :** Professionnels de santé, éducation, information des citoyens.

2.1. L'intégrer des sources des données patients hétérogènes

L'entrepôt de données permet avant tout d'intégrer les données cliniques hétérogènes depuis leur source pour permettre des analyses domaines. Il suffit de décrire les aspects d'intégration et d'analyse, l'architecture décisionnelle pour enfin présenter l'entrepôt de données comme centre d'info utile à la décision dans SIH [MAR14].

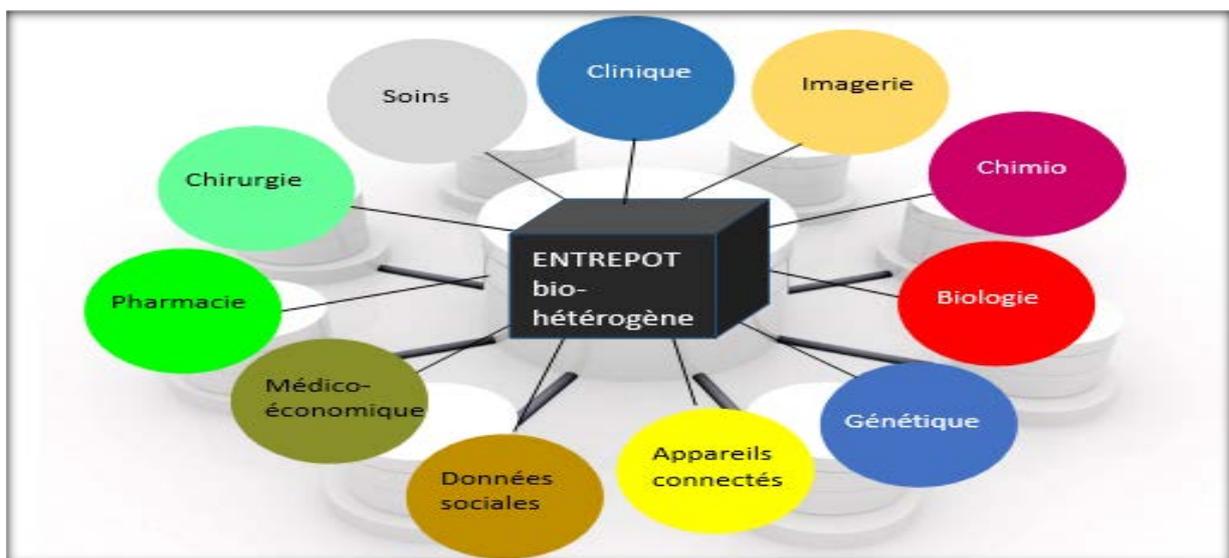


Figure 3.1 : données cliniques hétérogènes [W3]

3. Intégration des données de prescription dans un entrepôt de données biomédicales

LTSI Laboratoire Traitement du Signal et de l'Image avec CHU Rennes de France [LTC12] Ils ont intégré la prescription avec un moteur de recherche Roogle (système de recherche d'information au sein du dossier d'imagerie médicale).

Ils ont utilisé une base de données médicamenteuse indépendante et une définition standardisée des prescriptions et plus généralement du circuit de médicament, Et ils ont implémenté l'intégration à la recherche de prescription est possible.

Ils ont élaboré un schéma XML standard pour la prescription de l'intégration :

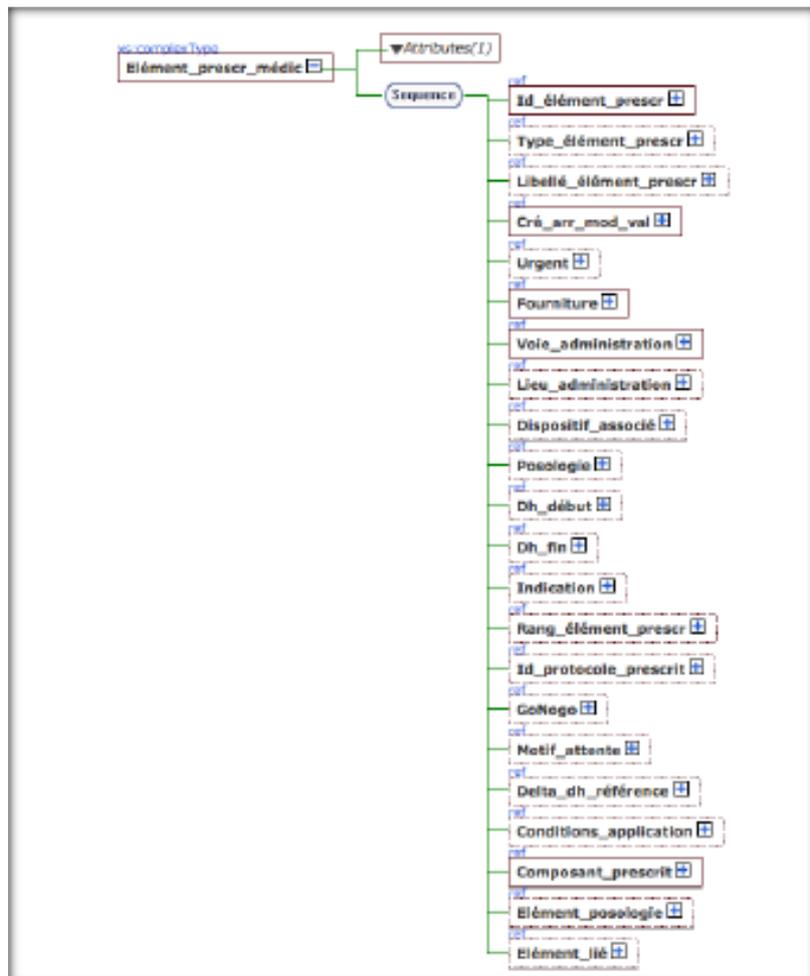


Figure 3.2 : Schéma XML d'un élément de prescription [LTC12]

4. Entreposage de données complexes pour la médecine personnalisée

Avec l'usage généralisé des nouvelles technologies, le domaine de la santé connaît des évolutions importantes qui laissent imaginer la possibilité d'un suivi personnalisé des patients tout au long de leur vie, de traitements pré symptomatiques ou encore d'analyses diverses concernant des populations de patients.

Jérôme Darmont [JER08] Il a présenté un modèle d'entrepôt de données complexes relatives à la santé d'athlètes de haut niveau. Et il a adopté une architecture en bus décisionnel. L'entrepôt de données est composé d'un ensemble de dimensions conformes et de définitions de faits.

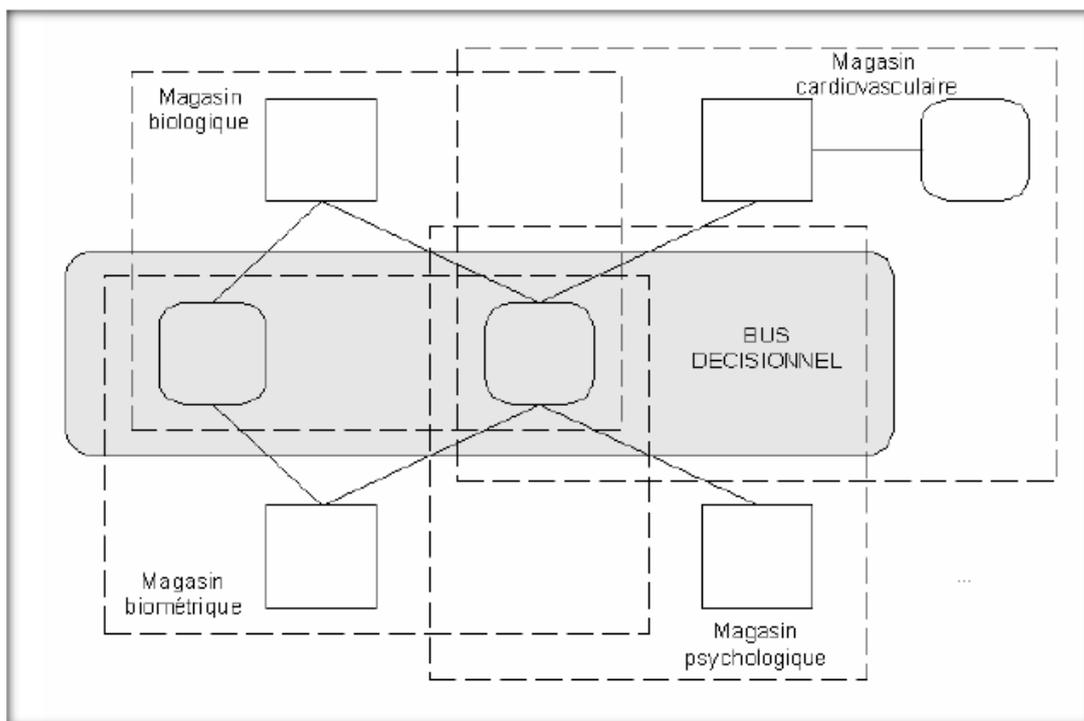


Figure 3.3 : Architecture globale de l'entrepôt de données [JER08]

Les carrés droits symbolisent les tables de faits, les carrés aux coins arrondis les dimensions, les pointillés entourent les différents magasins de données et le bus décisionnel est compris dans un rectangle arrondi grisé, Il est constitué par les dimensions communes à plusieurs magasins parmi elle [JER08]:

- les dimensions principales communes à tous les magasins sont les patients.
- les fournisseurs de données.
- le temps et l'analyse médicale.[JER08] il a conçu l'entrepôt de manière globale afin de :
- prendre en compte plusieurs spécialités médicales

Chapitre 03: Contribution : modèle d'intégration avec ETL et les services web.

- évolutive afin d'anticiper les avancées de la recherche médicale.
- permettre à la fois des analyses orientées patients.
- des études statistiques à large spectre.

5. Notre Proposition

Basé sur les travaux précédant, L'objectif principal de notre travail est de proposer et implémenter un modèle pour les SIH capable d'intégrer plusieurs sources hétérogènes au but de construire un entrepôt de données pour les informations des patients, Cette modèle prend en compte :

- Les sources sont hétérogènes.
- Nous avons utilisé l'approche ETL.
- Nous avons utilisé un service web pour résoudre le problème d'hétérogénéité.

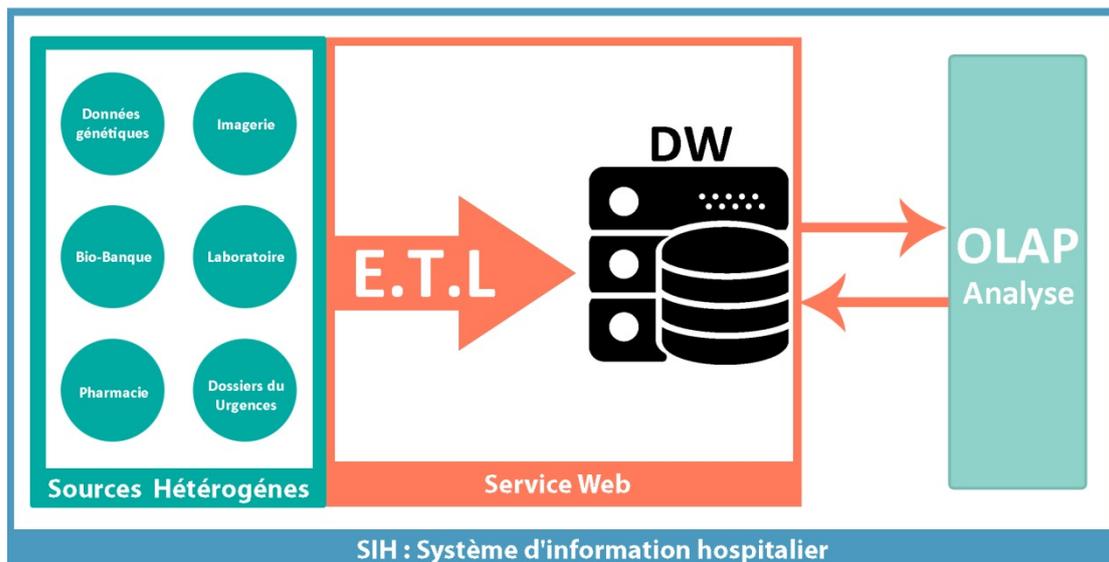


Figure 3.4 : système d'informatique hospitalier

5.1. Sources Hétérogènes

Plusieurs problèmes doivent être pris en compte pendant la conception de systèmes d'intégration. Ces problèmes résultent de l'hétérogénéité des données. Cette hétérogénéité est dû au fait que les sources de données ont été conçues indépendamment par des concepteurs différents, Cette hétérogénéité provient des choix différents qui sont faits pour représenter des faits du monde réel dans un format informatique.

Dans le SIH les sources sont hétérogènes par exemple Les données de la prescription ne sont pas les mêmes que les résultats du laboratoire, ou les résultats de l'imagerie et le dossier médical d'un patient.

5.2. Service Web

Pour résoudre le problème d'hétérogénéité, nous avons utilisé les services Web comme une interface pour les processus ETL :

- Extraction des données : En tant que première étape de l'intégration des données, le système extraira chaque collecte ou formulaire de données tel que : SQL, oracl, acceses, excel, ontologie avec des requetés SELECT ou SELECT ?, noud XML.
- Transformation : Comme deuxième étape du processus ETL, le système transformera toutes les données extraites par un schéma XML standard.
- Apres l'étape de transformation, le système charge les données vers l'entrepôt de données.

5.3. Data WareHouse

L'entrepôt de données recueille l'historique et les progrès du patient et toutes les informations sur le patient.

5.4. OLAP

OLAP est une méthode de stockage de données sous forme multidimensionnelle, généralement à des fins de génération de rapports, il peut stocker des informations comme le ferait une base de données traditionnelle, sa structure est radicalement différente. Historiquement, les bases de données sont conçues selon les exigences des systèmes informatiques qui les utilisent.

6. L'implémentation de modèle

Visual Studio .NET est un jeu complet d'outils de développement permettant de générer des applications Web ASP, des services Web XML, des applications bureautiques et des applications mobiles. Visual Basic .NET, Visual C++ .NET, Visual C# .NET et Visual J# .NET utilisent tous le même environnement de développement intégré (IDE, integrated development environment), qui leur permet de partager des outils et facilite la création de solutions faisant appel à plusieurs langages. Par ailleurs, ces langages permettent de mieux tirer parti des fonctionnalités du .NET Framework, qui fournit un accès à des technologies clés simplifiant le développement d'applications Web ASP et de services Web XML[Mic16].



Figure 3.5 : Visual Studio [Mic16]

Cette rubrique comporte des informations sur les éléments suivants [Mic16]:

- **Applications Smart Device :** L'environnement de développement intégré (IDE) de Visual Studio .NET comprend désormais des outils permettant de développer des applications pour les smart devices, comme le Pocket PC. À l'aide de ces outils et du .NET Compact Framework, un sous-ensemble du .NET Framework, vous pouvez créer, générer, déboguer et déployer des applications exécutées sur le .NET Compact Framework dans des assistants numériques personnels, des téléphones mobiles et d'autres périphériques limités en ressources.
- **Services Web XML :** Les services Web XML sont des applications qui peuvent recevoir des demandes et des données au format XML via le protocole HTTP. Il est possible d'accéder aux services Web XML à partir de n'importe quel langage, modèle de composant ou système d'exploitation dans la mesure où ces services ne sont pas tributaires d'une technologie de composant ou d'une convention d'appel d'objet particulières. Dans Visual

Studio .NET, vous pouvez rapidement créer et inclure des services Web XML en utilisant Visual Basic, Visual C#, JScript, les extensions managées pour C++ ou ATL Server.

6.1. Technologies XML

XML, qui signifie extensible Markup Language (traduisible par langage de balisage extensible, ou encore langage **a** balises extensible), est un langage constitué de balises, ou de tags.

Pour résumer, *XML* permet de répondre un tas de critères :

- *XML peut être utilisé sans difficulté sur Internet.*
- Il soutient une grande variété d'applications.
- Il facilite l'écriture de programmes traitant des documents XML.
- Les documents XML sont lisibles par l'homme et raisonnablement clairs.
- La conception du XML est rapidement exécutable.
- La facilité de créer des documents XML.
- Le langage peut s'auto-écrire.
- Le langage peut être étendu par lui-même : Les langages dérivés du XML sont définis en XML, ce qui permet de gérer tous les domaines d'application;
- Le contenu d'un document peut être lisible dans n'importe quel éditeur de texte et être compréhensible par une personne n'ayant aucune connaissance particulière;
- Le langage permet la description arborescente des données afin d'apporter par l'intermédiaire des arbres des structures de données effacées;
- Toute application munie d'un parseur de texte peut utiliser le document XML.

6.2. Sources hétérogènes:

Dans notre cas, nous allons utiliser les résultats d'un laboratoire pour un patient et les patients consultés.

CODP	NOM	PRENOM	DATE	UF	Plaquettes	VMP	PM	PE	PB	LY
1	NOM1	PRENOM1	01/06/2017	4001-HC HP P	22 G/L	8 FL	2,02 G/L	3,40%	0,90%	23,50%
2	NOM2	PRENOM2	02/06/2017	4001-HC HP P	23 G/L	9 FL	2,00 G/L	3,42%	0,90%	23,50%
3	NOM3	PRENOM3	03/06/2017	4001-HC HP P	24 G/L	10 FL	1,98 G/L	3,30%	0,90%	23,50%
4	NOM4	PRENOM4	04/06/2017	4001-HC HP P	25 G/L	11 FL	2,01 G/L	3,40%	0,90%	23,50%
5	NOM5	PRENOM5	05/06/2017	4001-HC HP P	26 G/L	12 FL	2,02 G/L	2,90%	0,90%	23,50%
6	NOM6	PRENOM6	06/06/2017	4001-HC HP P	27 G/L	13 FL	2,02 G/L	3,40%	0,90%	23,50%
7	NOM7	PRENOM7	07/06/2017	4001-HC HP P	28 G/L	14 FL	2,03 G/L	3,43%	0,90%	23,50%
8	NOM8	PRENOM8	08/06/2017	4001-HC HP P	29 G/L	15 FL	2,02 G/L	3,40%	0,90%	23,50%
9	NOM9	PRENOM9	09/06/2017	4001-HC HP P	30 G/L	16 FL	2,02 G/L	3,40%	0,90%	23,50%
10	NOM10	PRENOM10	10/06/2017	4001-HC HP P	31 G/L	17 FL	1,98 G/L	3,40%	0,90%	23,50%
11	NOM11	PRENOM11	11/06/2017	4001-HC HP P	32 G/L	18 FL	2,02 G/L	2,90%	0,90%	23,50%
12	NOM12	PRENOM12	12/06/2017	4001-HC HP P	33 G/L	19 FL	2,03 G/L	3,40%	0,90%	23,50%
13	NOM13	PRENOM13	13/06/2017	4001-HC HP P	34 G/L	20 FL	2,00 G/L	2,98%	0,90%	23,50%
14	NOM14	PRENOM14	14/06/2017	4001-HC HP P	35 G/L	21 FL	2,02 G/L	3,40%	0,90%	23,50%
15	NOM15	PRENOM15	15/06/2017	4001-HC HP P	36 G/L	22 FL	2,03 G/L	3,40%	0,90%	23,50%
16	NOM16	PRENOM16	16/06/2017	4001-HC HP P	37 G/L	23 FL	2,02 G/L	3,40%	0,90%	23,50%
17	NOM17	PRENOM17	17/06/2017	4001-HC HP P	38 G/L	24 FL	1,98 G/L	3,40%	0,90%	23,50%
18	NOM18	PRENOM18	18/06/2017	4001-HC HP P	39 G/L	25 FL	2,02 G/L	3,40%	0,90%	23,50%
19	NOM19	PRENOM19	19/06/2017	4001-HC HP P	40 G/L	26 FL	2,03 G/L	2,98%	0,90%	23,50%

Figure 3.6 : Résultats du labo type Excel

npat	ATC	lib	voie	debut	fin
1	DOLIPRANE	Antalgiques	Oral	8/01/2017	8/6/2017
1	VOLTARENE	arthroses	Oral	8/01/2017	8/6/2017
1	LEVOTHYROX	Endocrinologie	Oral	8/01/2017	8/6/2017
4	DOLIPRANE	Antalgiques	Oral	10/03/2017	9/9/2017
5	LEVOTHYROX	Endocrinologie	Oral	8/01/2017	8/6/2017
6	LEVOTHYROX	Endocrinologie	Oral	8/01/2017	8/6/2017
1	MAGNE B6	Métabolisme	IV	8/01/2017	8/6/2017
3	DAFLON	Cardiologie	IV	3/5/2017	2/11/2017
5	ELUDRIL	Stomatologie	Oral	2/6/2017	1/12/2017

Figure 3.7 : liste de préinscription type Access

Nom	poid	taille	maladie	date
NOM1	65	173	Amylose	8/1/2017
NOM2	72	172	Appendicite	3/2/2017
NOM3	60	170	Acné	3/5/2017
NOM4	81	174	Aérophagie	10/03/2017
NOM5	81	171	Albinisme	8/01/2017
NOM6	75	174	Brucellose	8/01/2017
▶*	NULL	NULL	NULL	NULL

Figure 3.8 : liste des consultations type SQL

6.3. Service Web et ETL

✓ Application d'extraction

Pour l'extraction de données, nous avons développé une application qui peut se connecter à différentes sources de données.

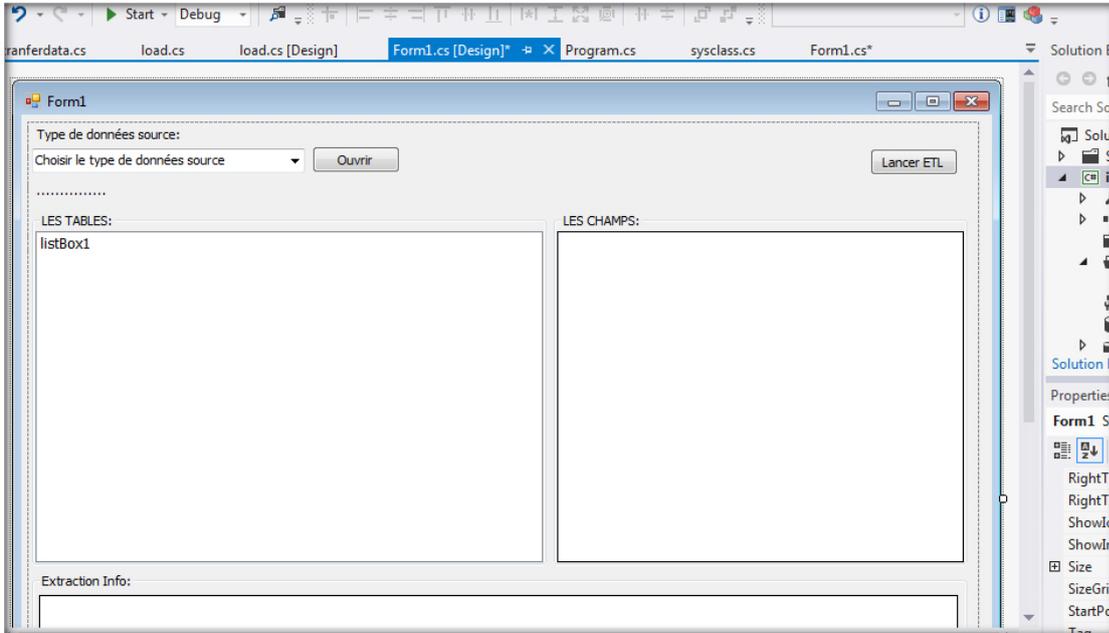


Figure 3.9 : Interface d'application d'extraction

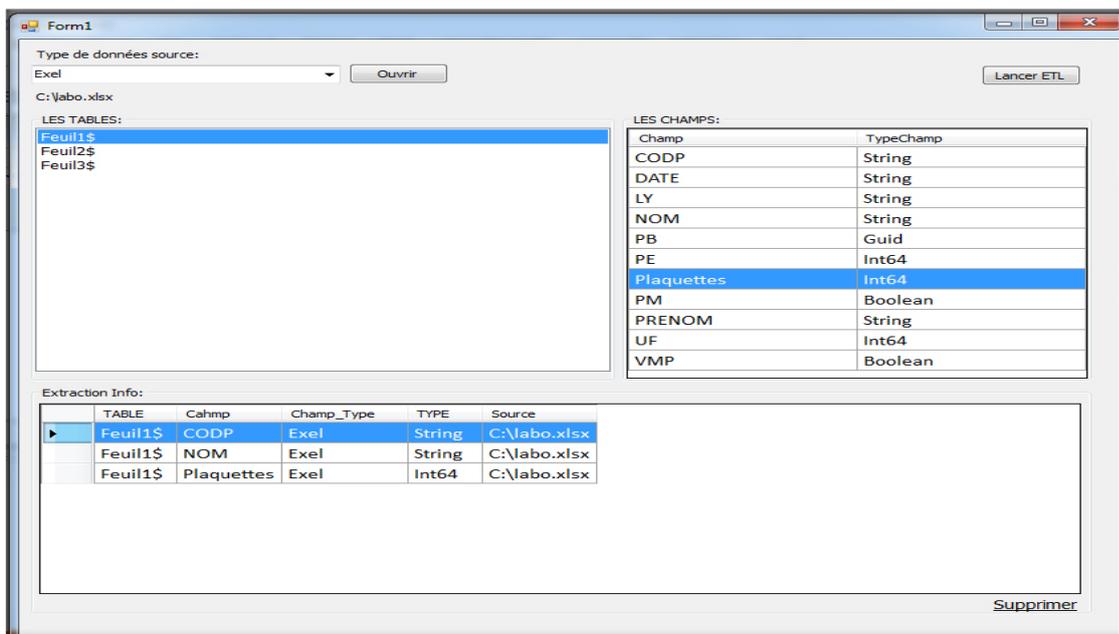


Figure3.10 : Interface d'application d'extraction en exécution

Chapitre 03: Contribution : modèle d'intégration avec ETL et les services web.

Dans la figure 3.10, nous avons sélectionné une source de données de type Excel, et nous avons sélectionné la table et les attributs à extraire, dans notre exemple Nous avons sélectionné les attributs suivants : CODP, nom et plaquettes.

✓ Service web

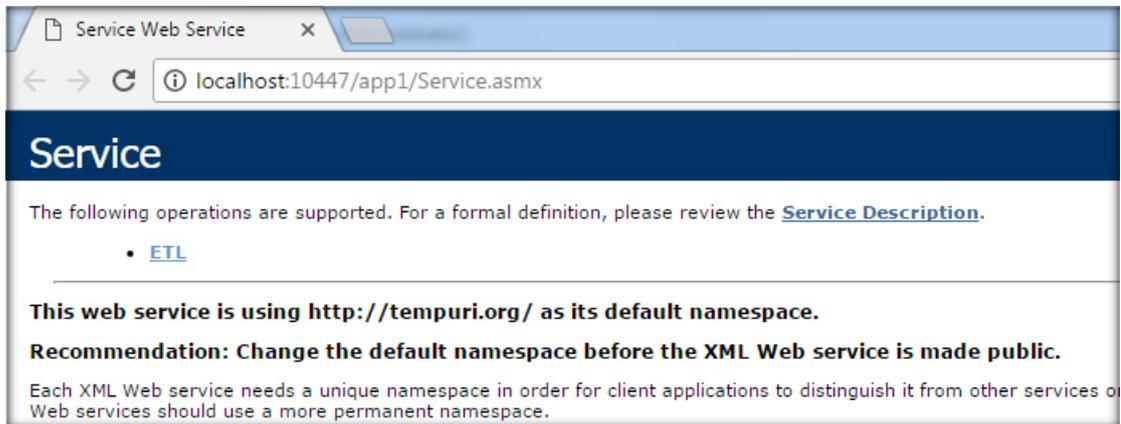


Figure 3.11 : l'interface de ETL au niveau de service web

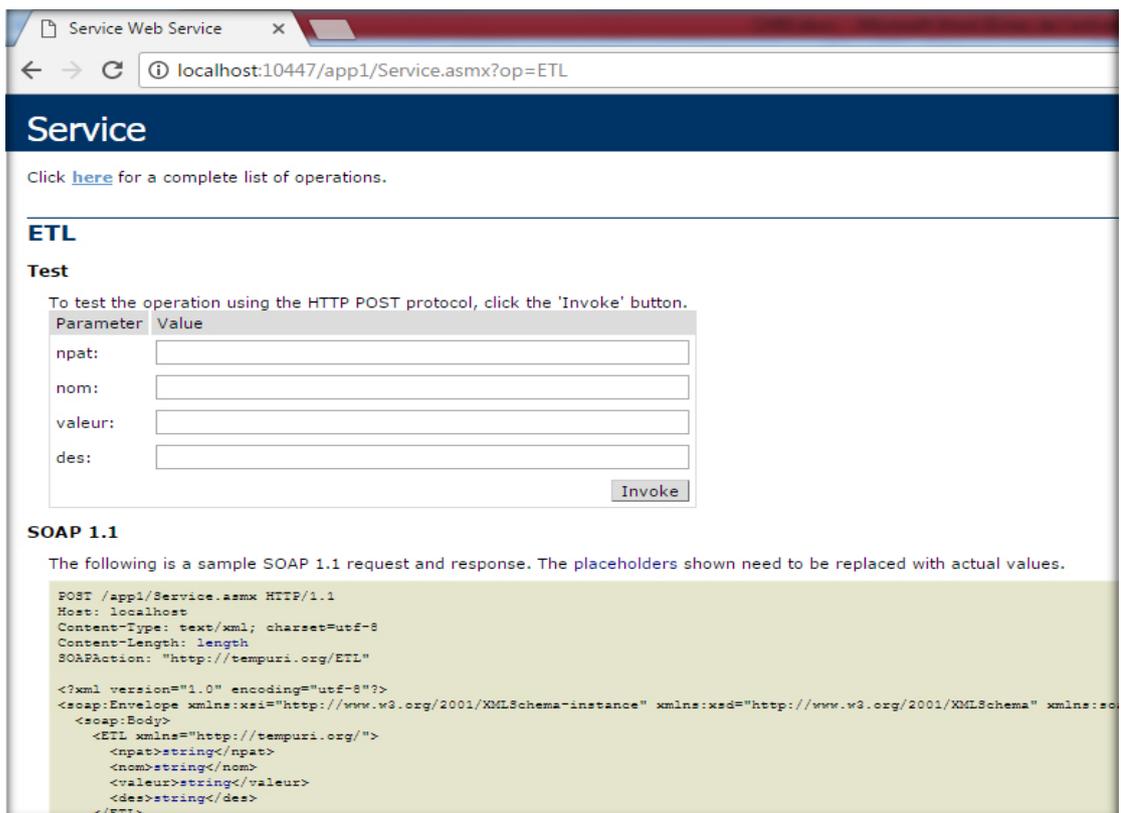


Figure 3.12 : la méthode ETL au niveau de service web

Chapitre 03: Contribution : modèle d'intégration avec ETL et les services web.

La méthode a des paramètres tels que : npat(N° patient) , nom (nom du patient) valeur(données extracté) et des (type de source de donnée),Chacun manipule la valeur extraite et le schéma xml pour transformer les données extraites, après l'extraction et la transformation, la méthode adddata() appelle un autre mode pour charger les données vers le DW.

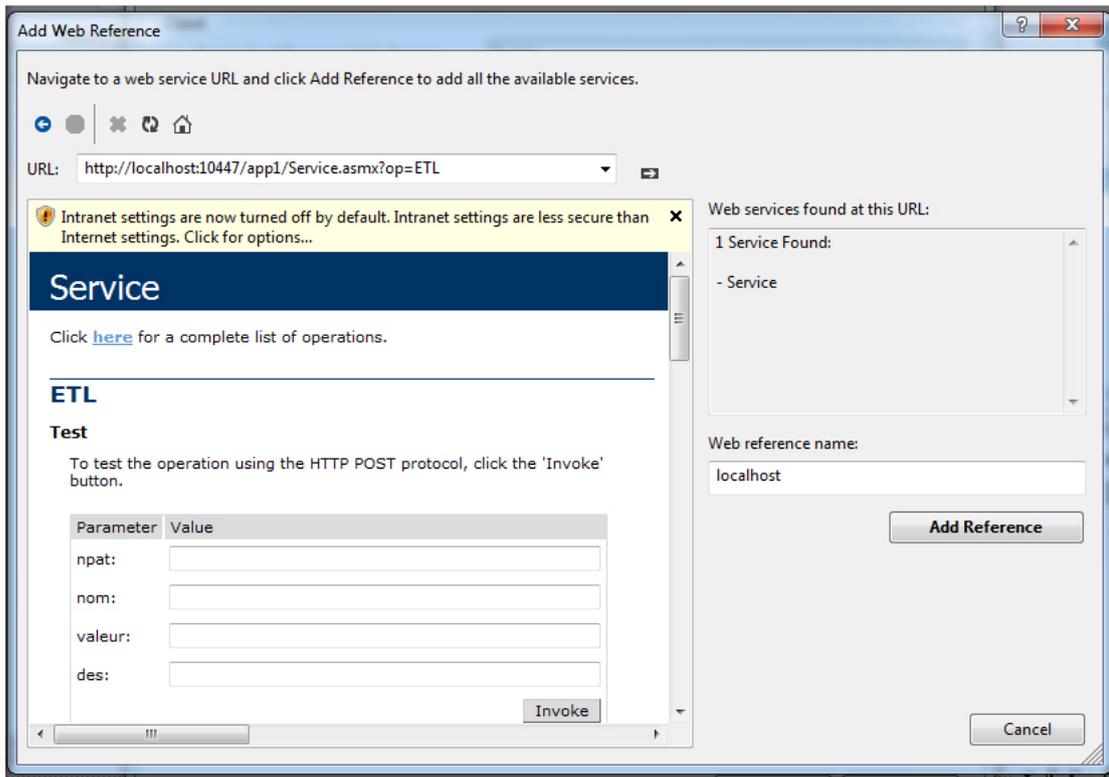


Figure 3.13 : connexion entre l'application d'extraction et service web

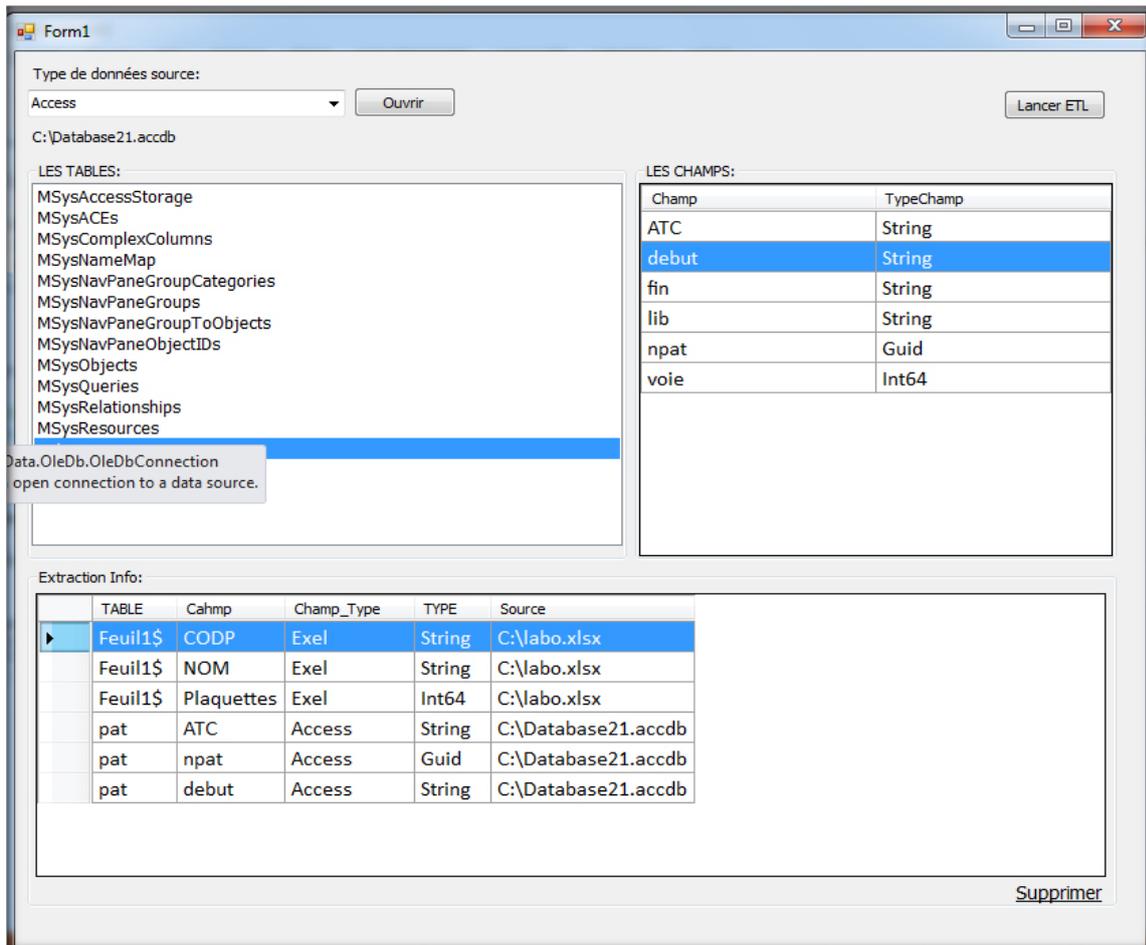


Figure 3.14 : la sélection des champs et sources

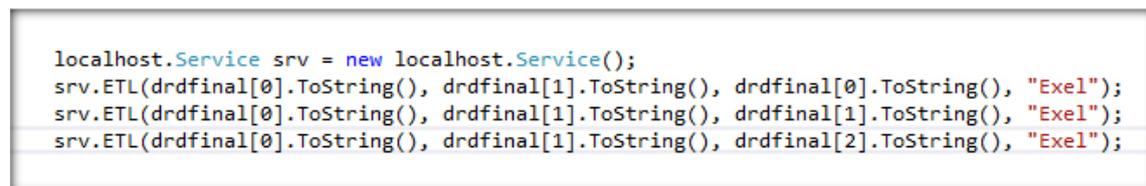
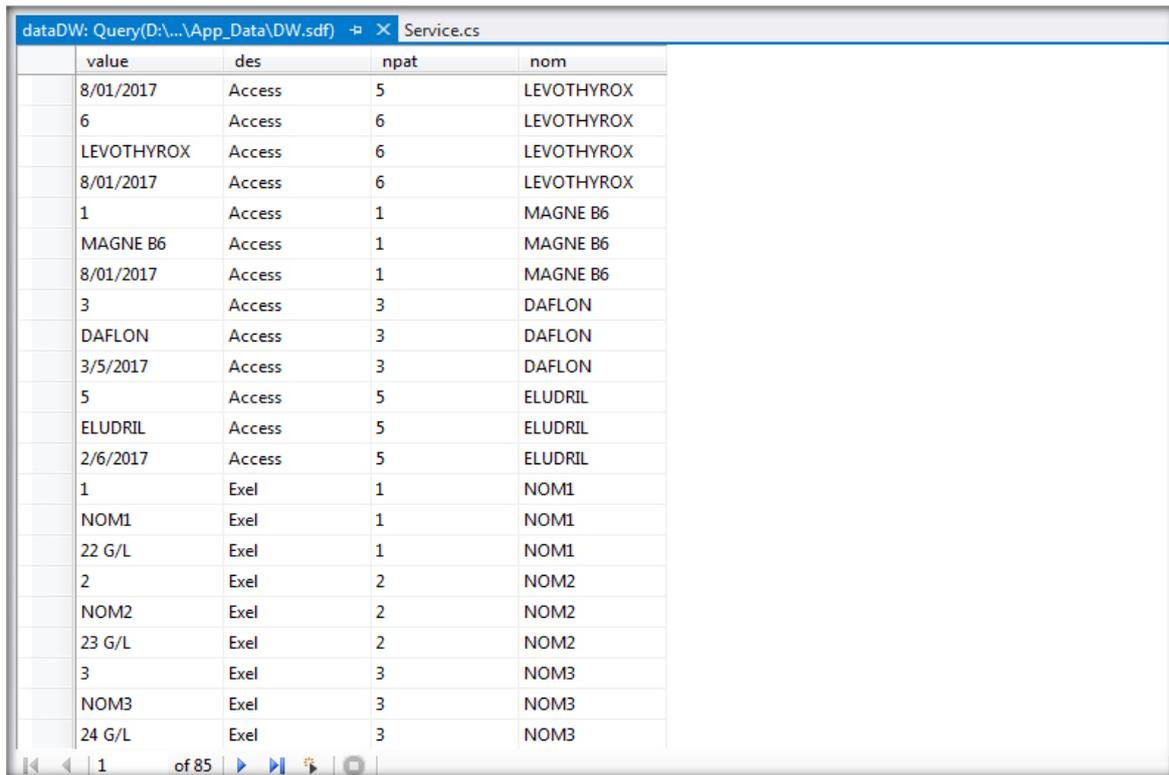


Figure 3.15 : l'appelle de méthode ETL au niveau d'application d'extraction

Chapitre 03: Contribution : modèle d'intégration avec ETL et les services web.



value	des	npat	nom
8/01/2017	Access	5	LEVOTHYROX
6	Access	6	LEVOTHYROX
LEVOTHYROX	Access	6	LEVOTHYROX
8/01/2017	Access	6	LEVOTHYROX
1	Access	1	MAGNE B6
MAGNE B6	Access	1	MAGNE B6
8/01/2017	Access	1	MAGNE B6
3	Access	3	DAFLON
DAFLON	Access	3	DAFLON
3/5/2017	Access	3	DAFLON
5	Access	5	ELUDRIL
ELUDRIL	Access	5	ELUDRIL
2/6/2017	Access	5	ELUDRIL
1	Exel	1	NOM1
NOM1	Exel	1	NOM1
22 G/L	Exel	1	NOM1
2	Exel	2	NOM2
NOM2	Exel	2	NOM2
23 G/L	Exel	2	NOM2
3	Exel	3	NOM3
NOM3	Exel	3	NOM3
24 G/L	Exel	3	NOM3

Figure 3.16 : le Data Warehouse.

Conclusion

Pour résoudre les problèmes majeurs d'hétérogénéité entre l'intégration des données massive nous avons essayé de proposer un modèle d'intégration des données dans les SIH, avec les technologies des services web et XML.

Conclusion Générale

Dans cette étude, nous avons entamé un sujet de haut importance et d'actualité qu'est la médecine personnalisée qu'est le Big Data.

Après avoir détaillé ces deux axes et dégagé les problèmes liés, nous avons constaté que les travaux de recherche sur la médecine personnalisée et Big Data sont rares et nous n'avons pas constaté des approches proposés. Ainsi nous proposons dans une première étape une approche théorique qui tient compte de tous les acteurs liés au patient. Nous avons également essayé de résoudre le problème d'hétérogénéité en utilisant les services Web et XML.

Bibliographies:

[AI,AB15]	Abdelfettah I, Azedine B, " Aspects techniques d'un modèle de fouille de données Cloud basé... , 2015.
[AB15]	A Bouzidi, " Systèmes d'aide à la décision médicale ", 2015.
[AG17]	Agence Gimmik , " livret-l'imagerie Médicale - CEA ", 2017.
[AG15]	Alexandre Galopin, " Modélisation ontologique des recommandations de pratique clinique pour une aide à la décision `a niveaux d'abstraction variables" , 2015.
[AK, IL13]	Angeline k, Insa Lyon, " Memoire Online - Big data (rapport de stage) – Angeline KONE ", 2013
[DP, FM05]	Degoulet P, Fieschi M, " Traitement de l'information médicale: méthodes et applications ", Ed Masson, 2005.
[BES09]	Berner ES, "Clinical decision support systems: State of the Art". AHRQ Publication N°. 09 - 0069 - EF. Rockville, Maryland: Agency for Healthcare Research and Quality, June 2009.
[BL16]	Basten L, " Le Big Data, l'IA et le Machine Learning transforment les soins de santé ", 2016.
[BM14]	Bernard Marr, " Big Data: The 5 Vs Everyone Must Know Bernard Marr Pulse ... ", 2014
[BEN15]	BENOSMAN Amina, " Médiation sémantique de données basée sur les services web : Application dans le domaine médical ", 2015
[BS,JB]	Brigitte Seroussi, Jacques Bouaud, " Les systèmes experts pour l'aide à la décision médicale.
[BS,JB15]	Brigitte S´eroussi, Jacques Bouaud, " Systèmes informatiques d'aide à la décision en médecine : panorama des approches utilisant les données et les connaissances ", 2015.
[CH11]	S. Chafki, C. Desrosiers ETS Département de génie logiciel et des TI MTI820 Hiver 2011.
[DAR03]	DARMOUNI, " informatique de santé, Sciences et Technologies de l'information se de la communication ",2003
[DJ,GS04]	Dean, J., Ghemawat, S, "Mapreduce: Simplified data processing on large clusters". In: 6th Symposium on Operating System Design and Implementation (OSDI 2004), San Francisco, California, USA, December 6-8, 2004.
[DP, FM05]	Degoulet P, Fieschi M, " Traitement de l'information médicale: méthodes et applications ", Ed Masson, 2005.
[DSB99]	B. Dinter, C. Sapia, M. Blaschka, and G. Hofling, " OLAP Market and Research

	: Initiating the Cooperation. Computer Science and Information Management, 2(3)", 1999 .
[FAV07]	C. Favre , " Evolution de schémas dans les entrepôts de données : mise a jour de hiérarchies de dimension pour la personnalisation des analyses. Interface homme-machine [cs.HC]. Université Lumière - Lyon II", 2007 .
[EMA11]	EMA " Personalised medicine towards the market and patients: the approval Process", European Medicines Agency's perspective, mai 2011 .
[FDA13]	FDA , " Personalized medicine. FDA's unique role and responsibilities in personalized Medicine", Octobre 2013 .
[ISM05]	Ingelman. Sundberg M , " Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity". Pharmacogenomics J, 2005, 5, 6 –13",2005
[IWH94]	Inmon W.H, Hackarton , "RD: Using the Data Warehouse. John Wiley & Sons", 1994 .
[JAM94]	Jambu, M. , "Introduction au Data Mining. Analyse intelligente des données, Paris Eyrolles, Paris, 1999 , 144p", 1999 .
[JMM09]	Jean-Marc Manach , " Informatique médicale : patient virtuel, usine à gaz et pompe à fric ", 2009
[JER08]	Jérôme Darmont , Entreposage de données complexes pour la médecine d'anticipation personnalisée : Université de Lyon, Lyon 2, France 2008 .
[KNA13]	Kats Nelson A , " Momentum grows to make 'personalized' medicine more 'precise'. Nature Med 2013; 19(3): 249",2013 .
[KB,KM12]	Kevin Brenkel, Kulani Makhubele "Expert System", Medical Advisor System, http://pubs.cs.uct.ac.za/honsproj/cgi , 2012 .
[LHE95]	Lesca H., Lesca E , " Gestion de l'information . qualité de l'information et performances de l'entreprise, Editions Litec", 1995.
[LF12]	L Foucan , "Introduction à l'Informatique Médicale", 2012 ..
[LOU10]	Louisa Demmou essai présenté au Centre de formation en technologie de l'information (CeFTI) en vue de l'obtention du grade de Maître en Génie Logiciel MGL, 2010 .
[MB14]	Mickael Baron , " Introduction à Apache Hadoop : généralités sur HDFS et MapReduce", 2014
[MAR14]	Pr Marc CUGGIA , " Equipe projet données massives en santé (Inserm UMR 1099 LTSI) Département d'information médicale", 2014 .

[MC12]	M.CORINUS, T.Derey, J.Marguerie, W.Techer, N.Vic , "Rapport d'étude sur le Big Data", SRS Day, 6p, 2012 .
[NS, HS15]	Nitin Sawant, Himanshu Shah , " Big Data Application Architecture Q&A - A Problem - Solution Nitin ... " , 2015
[RW,DA95]	Rosenberg W, Donald A. ," Evidence based medicine: an approach to clinical problem-solving. <i>BMJ</i> , 1995, 310, 1122-1126" , 1995 .
[SAJ15]	Scheen AJ , "Editorial. Médecine conventionnelle, médecine factuelle, médecine personnalisée: trois approches complémentaires. <i>Rev Med Liège</i> , 2015, 70, 221-224" , 2015
[SEG14]	SEGHIRI Narimane ,"détection et extraction des information temporelles", 2014.
[SB12]	Shad Burks , "l'importance de l'informatique dans le secteur de la santé", 2012 .
[SEH 87]	Shortliffe EH , "Computer programs to support clinical decision making", <i>JAMA</i> , 258(1), 1987
[W1]	Big Data : comparatif des tous les articles... http://www.journaldunet.com/solution/analytics/bigdata/ .
[W2]	Le Big Data , "Définition : Qu'est-ce que le Big Data ? - Le BigData. Fr" , http://www.lebigdata.fr/definition-big-data , 2017 .
[W3]	Ngiambus Marcus , "Marc NGIAMBA" , http://ngiambus.over-blog.com/2014/11/qu-est-ce-que-le-big-data.html , 2014
[W4]	Des métiers pour l'informatique, http://www.awt.be/web/inf/index.aspx?page=inf,fr,200,010,001 " , 2015
[W5]	Shade Burks, Image ordinateur , " http://imageordinateur.blogspot.com/2012/11/importance-de-l-dans-le-secteur-de-la.html " , 2012
[W6]	Médecine personnalisée, Psychomédia, 3mai 2016 http://www.psychomedia.qc.ca/lexique/definition/medecine-de-precision
[LTC12]	"Intégration des données de prescription dans un entrepôt de données biomédicales Laboratoire Traitement du Signal et de l'Image (LTSI) INSERM : U1099 – Université de Rennes et CHU Rennes, Pôle Pharmacologie et Investigation cliniques, Unité « Fouille de données », F35033 Rennes", France 2012 .