



REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE IBN KHALDOUN - TIARET

MEMOIRE

Présenté à :

FACULTÉ MATHÉMATIQUES ET INFORMATIQUE
DÉPARTEMENT D'INFORMATIQUE

Pour l'obtention du diplôme de :

MASTER

Spécialité : Génie Informatique

Par :

TORCH Mohamed
MAASKRI Ali

Sur le thème

Gestion temps réel d'une base de données d'images médicales

Soutenu publiquement le 18/09/2022, à Tiaret devant le jury composé de :

BOUDAA Boudjema	MCA	Université Tiaret	Président
BELARBI Mustapha	Pr	Université Tiaret	Encadreur
CHENNINE Abdekader	MCB	Université Tiaret	Examineur

2021-2022

Remerciements

Tout d'abord, nous voudrions remercier « **ALLAH** » Tout-Puissant et Miséricordieux. Créateur de tout l'univers qui nous a donné la santé, la force, le courage et la volonté de mener à bien cet humble travail de recherche.

Nous voudrions remercier toutes les personnes ayant soutenu et encadré ce travail :

En premier lieu, nous remercions sincèrement et très chaleureusement, nos encadreurs, Mr. **BELARBI Mustapha**, Professeur à l'université Ibn Khaldoun de Tiaret, pour tout, que ce soit en raison de leur disponibilité ou de leur principale contribution à la direction des travaux de recherche et de leurs idées originales et éclairantes qui ont influencé le contenu de ce mémoire de Master et ouvert la voie à son achèvement.

Nous remercions également Mr. **BELABBES Kadda** pour l'aide qu'il nous a apporté et les précieux conseils qui ont abouti à ce travail.

Nous tenons à remercier Mr. **BELFEDHAL Salah-Eddine**, Etudiant Automatismes à l'université Ibn Khaldoun de Tiaret.

Nous tenons également à remercier Mr. **BOUDAA Boudjemaa**, Maître de Conférences classe A à l'université Ibn Khaldoun de Tiaret, pour l'intérêt qu'elle a porté à ce travail en nous honorant par sa présence dans le jury.

Nous tenons également à remercier Mr. **CHENNINE Aek**, Maître de Conférences classe A à l'université Ibn Khaldoun de Tiaret, pour l'intérêt qu'elle a porté à ce travail en nous honorant par sa présence dans le jury.

Enfin, nos sincères remerciements vont à tous ceux qui ont contribué à la réalisation de ce travail, de près ou de loin.

Dédicace

Je dédie ce Travail à mes parents et à mes frères et sœurs.

Je voudrais aussi remercier ma chère épouse Oum

Anes et mon enfant Anes

Et à tous les personnels de la

commune de Sougueur.

TORCH Mohamed

Dédicace

Je dédie ce fruit à mes parents et à mes frères et sœurs.

Je voudrais aussi remercier ma chère épouse

et ma Fille Rahaf Maram

Et à tous mes amis.

MAASKRI Ali

الملخص

معالجة الصور تلعب دورًا مهمًا اليوم في العديد من المجالات، لا سيما في التصوير الطبي. يعد تسجيل الصور تقنية أساسية في معالجة الصور الطبية وتحليلها، والتي تهدف إلى متابعة حالات المرضى، لذا فالصورة مرجعية من أجل استخراج أصغر الاختلافات التي يمكن استخدامها لاحقًا من قبل المتخصصين في هذا المجال. في هذا العمل ، نقدم نهجًا جديدًا لتسجيل الصور بطريقة مباشرة في قاعدة بيانات مشتركة وكبيرة، وذلك للاستفادة منها في حالات مشابهة للمصابين بالسرطان.

Résumé

Le traitement d'images joue aujourd'hui un rôle important dans de nombreux domaines, notamment en imagerie médicale. L'enregistrement d'images est une technique essentielle dans le traitement et l'analyse d'images médicales, qui vise à suivre l'état des patients, l'image est donc une référence afin d'extraire les plus petites différences qui peuvent être utilisées ultérieurement par les professionnels de ce domaine. Dans ce travail, nous présentons une nouvelle approche pour enregistrer directement des images dans une grande base de données commune, pour une utilisation dans des cas similaires de patients atteints de cancer.

Abstract

Image processing plays an important role today in many areas, particularly in medical imaging. Image registration is an essential technique in medical image processing and analysis, which aims to follow up on patients' conditions, so the image is a reference in order to extract the smallest differences that can be used later by professionals in this field. In this work, we present a novel approach to directly register images in a large common database, for utilization in similar cases of patients with cancer.

Liste des tableaux

Tableau 1 Présentation des différents de tumeur en fonction du tissu d'origine.....	49
Tableau 2 Principaux types histologiques des cancer de la thyroïde[50].....	52
Tableau 3: Nouveaux cas de cancers enregistrés en Algérie de 2015 à 2017.....	57
Tableau 4 : Nouveaux cas de cancers survenus en Algérie en 2017.....	57
Tableau 5 : Exemple de séquence de création des diagrammes.....	80

Liste des figures

Figure 1: Figure Cancer de la prostate	50
Figure 2: cancer du poumon.....	51
Figure 3: Cancer de la thyroïde	52
Figure 4: Les 03 Réseaux Régionaux des Registres de Cancer.....	56
Figure 5 : Shimadzu Europe-UV-1650PC.....	67
Figure 6 : Spectrophotomètre UV-VIS V-630	70
Figure 7 : Fonction Spectrophotomètre UV-VIS V-630	71
Figure 8 : Affichage Spectrophotomètre UV-VIS V-630 [60].....	72
Figure 9 : Rigaku MiniFlex 600.....	72
Figure 10: Installation de réseau (Schéma tracer par Cisco Packet Tracer).....	75
Figure 11: Modélisation d'UML	81
Figure 12 : Diagramme cas d'utilisation.....	82
Figure 13 : Diagramme de classe (Qui présenter les classes d'application)	83
Figure 14: Diagramme de séquence (Authentification).....	84
Figure 15: Diagramme de séquence (Création d'utilisateur).....	85
Figure 16 : diagramme de séquence (création de rendez-vous)	86
Figure 17 : Interface entrée d'application par administrateur	87
Figure 18 : Interface après entrée l'administrateur.....	87
Figure 19: Interface pour crée nouveau médecin, Employée et administrateur	88
Figure 20: Interface pour entée un nouveau patient (malade)	88
Figure 21: Interface crée médecin.....	89
Figure 22: Interface entée par un médecin.....	89
Figure 23: Interface autorisé au médecin.....	90
Figure 24: Interface entrée par un utilisateur	90
Figure 25: Entée un nouveau patient par l'employeur	91
Figure 26: Créé un nouveau médecin par l'employeur.....	91
Figure 27: Créé un rendez-vous par un employeur	92

Contenu

Introduction générale.....	11
CHAPITRE I Big Data et Hadoop	11
1. Big Data	12
1.1. Introduction :	12
1.1.1 Définition de Big Data :	12
1.1.2. Caractéristiques du Big Data :	13
1.1.3.1 Vitesse	13
1.1.3.2 Volume.....	13
1.1.3.3 Variété	13
1.1.3 Architecture Big Data :	13
1.1.4 Avantages de l'architecture Big Data :	14
1.1.5 Domaines d'utilisation du Big Data :	14
1.1.6 Les principales technologies de Big Data :	15
1.1.7 Bases données NoSQL :	15
1.1.7.1 Caractéristiques NoSQL	16
1.1.7.2 Les types des bases NoSQL	16
1.1.7.3 Les principales bases de données NoSQL :	17
2.L'Annotation des images :	17
3.Introduction d'Hadoop :	18
3.1 Le système de fichier distribué d'Hadoop HDFS :	19
3.3.2 Les HDFS définit deux types de nœuds :	19
3.3.2.1Le nœud principal ou NameNode.....	19
3.3.2.2Le nœud de données ou DataNode	20
3.3.2.3Secondary NameNode	20
3.3.3 Lecture d'un fichier HDFS :	20
3.3.4Ecriture dans un fichier HDFS	21
3.4 Ecosystèmes d'Hadoop :	21
3.5 Mise en œuvre et l'installation de Hadoop :	21
3.5.1 L'environnement de travail :	22
3.5.1.1 Virtual Box.....	22
3.5.1.2 Ubuntu	22
3.5.1.3 Choix d'Ubuntu 20.04:.....	23
3.5.1 Les étapes d'installation et configuration de l'environnement de travail.....	23

CHAPITRE II Etude de cas Le Cancer dans l'Algérie	36
Introduction :	45
2. Généralité sur le cancer :	46
2.1. Définition	46
2.2. Epidémiologie du cancer.....	47
2.3. Les facteurs de risque de cancer	47
2.4. Cancérogénèse	47
2.5. Classification des cancers.....	48
2.5.1. La classification TNM des cancers	49
2.6.1. Cancer du sein	49
2.6.2. Cancer de la prostate	50
2.6.3. Cancer du poumon	51
2.6.4. Cancer de la thyroïde.....	52
2.7.1. Examens biologique.....	53
2.7.2. Marqueurs tumoraux	53
2.7.3. Examens d'imageries médicales	53
2.7.4. Examens histologique.....	53
2.8. Le cancer en Algérie.....	54
2.8.1. Description des réseaux de registres de cancer.....	55
2.9. Traitement	57
2.9.1 Chirurgie.....	57
2.9.2. Radiothérapie	57
2.9.3. Traitement médicamenteuses.....	58
2.10 Plan national cancer 2015-2019	59
2.10.1 Objectif principal du plan	59
3. Propositions de mesures	63
3.1- L'amélioration de l'accueil du patient.....	63
3.2- La disponibilité des moyens thérapeutiques.....	63
3.3- La rapidité du diagnostic.....	63
3.4- La mise en application effective des textes réglementaires	63
3.5- Le rôle capital et l'intérêt de l'intersectorialité.....	63
3.6- L'adaptation de l'enseignement aux nouveaux métiers.....	63
Conclusion.....	64
CHAPITRE III Conception d'application.....	65

Introduction	67
1. Spécification informelle du banc de test anapath	67
1.1 Aspect matériel	67
2. Spectrophotomètre UV-VIS V-630 [59]	70
3. Diffractomètre à rayons X (Rigaku MiniFlex 600)	72
4. Structure de réseau	75
5. Conception	75
5-1. UML	75
5.1.1 Utilisation.....	76
5.1.2 Formalisme	76
5.1.3 Vues d'UML.....	77
5.1.4 Diagrammes UML.....	77
5.1.5 Modèles d'éléments.....	79
5.1.5 Modèles d'éléments de type commun	79
5.1.6 Modèles d'éléments de type relation	80
5.1.7 Autres modèles d'éléments.....	80
5.2.8 Normalisation et certification.....	80
5.2.9 Logiciels de modélisation UML	81
5.3. Diagramme de cas d'utilisation.....	82
5.4. Diagramme de classe.....	83
5.5. Diagramme de séquence	84
6. l'Application	87
Conclusion	93

Introduction

générale

Introduction générale

Introduction générale

L'imagerie médicale est certainement l'un des domaines de la médecine qui a le plus progressé ces deux dernières décennies. La première application d'imagerie médicale date de 1895 avec l'utilisation des rayons X. C'est le début de la radiographie. Ensuite, les techniques s'améliorent et se diversifient avec l'arrivée de la scintigraphie, du scanner, de l'échographie puis de l'IRM. Ces récentes découvertes permettent non seulement un meilleur diagnostic mais offrent aussi de nouveaux espoirs de traitement pour de nombreuses maladies comme le cancer, l'épilepsie... l'identification précise de la lésion facilite déjà le recours à la chirurgie, seule solution thérapeutique pour certains malades. De telles techniques permettent également de mieux comprendre le fonctionnement de certains organes encore mystérieux, comme le cerveau.

Outre la partie introductive et la conclusion générale, le travail est organisé en trois chapitres.

Le chapitre 1 : a pour titre «Big Data et Annotation Vers Hadoop», ce chapitre présente l'installation et le fonctionnement de Hadoop.

Le chapitre 2 : a pour titre «Cancer dans l'Algérie», ce chapitre est consacré à la présentation de la propagation du cancer en Algérie.

Le chapitre 3 : nous présentons dans ce chapitre en détail la conception de l'application et son fonctionnement.

Nous terminons avec une **conclusion générale** : dans cette partie, nous récapitulons notre contribution et proposerons des perspectives sur la base des travaux effectués dans notre travail.

Problématique :

Le processus de traitement des images médicales est effectué dans plusieurs équipements distincts, ce qui nécessite du personnel du laboratoire pour traiter ces images par les médecins,

Objectif :

L'objectif principal du projet est :

1. Réaliser un programme d'affichage des images médicales
2. Créer des rapports médicaux et consulter ces rapports de plusieurs médecins

CHAPITRE I Big Data et Hadoop

1. Big Data

1.1. Introduction :

Des milliers d'entreprises, d'organisations et d'individus publient de grandes quantités d'informations sur le Web, et la charge de travail et la quantité de données pouvant être gérées ont augmenté rapidement car ces entreprises ont eu recours à

Entrepôt de données ou entrepôt de données pour l'analyse et le stockage des données. Typiquement, le Datawarehouse est centralisé sur un serveur rattaché à une baie de disques, solution difficilement évolutive (autre que la consommation électrique), sauf qu'il ne gère que les données structurées dans le SGBD.

Pour faire face à l'explosion de la quantité de données, nous parlons maintenant de pétaoctets (milliards d'octets) ou même de setaoctets (billion d'octets) et aussi lorsque nous travaillons avec de grandes quantités de données (images, texte, web, etc.) une nouvelle. Le domaine de la technologie a émergé : le Big Data, inventé par les grands géants du web tels que Yahoo, Google et Facebook, qui ont été les premiers à mettre en œuvre ce type de technologie.

Ce concept fournit une architecture distribuée et évolutive pour le traitement et le stockage des données. L'objectif principal de ce nouveau paradigme est d'améliorer les performances et d'augmenter l'exécution et le traitement des requêtes.

1.1.1 Définition de Big Data :

Le terme Big Data a été mentionnée pour la première fois par le cabinet d'études Gartner en 2008, mais la naissance de ce terme efficace a commencé en 2001 et a été mentionnée par le Meta Group.

Elle entend faire exploser la quantité de données (en termes de volume, de vitesse de production et de diversité) et proposer de nouvelles solutions pour gérer ce volume en termes de capacité de stockage et d'exploration, et plus récemment la possibilité d'analyser et d'utiliser ces données dans temps réel.

Big data est un terme anglais utilisé pour décrire des ensembles de données qui sont devenus si volumineux qu'il a été difficile de travailler avec des outils de gestion de base de données traditionnels. Il s'agit donc d'un ensemble de technologies, d'architectures, d'outils et de méthodes qui permettent aux organisations de capturer, traiter et analyser facilement de grandes quantités et contenus.

Obtenir des informations hétérogènes et variables et pertinentes à un prix abordable.

1.1.2. Caractéristiques du Big Data :

Le Big Data ("Big Data" en français) fédère une famille d'outils sont : la vitesse et le volume ainsi que la variété.

1.1.3.1 Vitesse : La vitesse décrit la fréquence à laquelle les données sont générées, récupérées et partagées. Les entreprises doivent comprendre la vitesse non seulement en termes de création de données, mais également en termes de traitement, d'analyse et de livraison à l'utilisateur conformément aux exigences des applications en temps réel.

1.1.3.2 Volume : Le Big Data est associé à une quantité déroutante de données, qui varie désormais de dizaines de téraoctets à plusieurs pétaoctets dans l'ensemble de données.

Les entreprises et tous les secteurs d'activité doivent trouver des moyens de gérer la quantité toujours croissante de données générées chaque jour. Les catalogues de plus de 10 millions de produits sont devenus la règle au lieu de l'exception.

Voici quelques photos pour illustrer ce phénomène :

- 90 % des données actuelles n'ont été créées qu'au cours des deux dernières années ;
- Twitter, par exemple, génère 7 To de données par jour.

1.1.3.3 Variété : Le développement de données différentes est le résultat de nouvelles données multi structurales et de l'extension de types de données provenant de différentes sources.

Hétérogène. Aujourd'hui, on trouve des capteurs d'informations dans les appareils électroménagers, les trains, les voitures ou les avions, qui produisent de nombreux types d'informations différents.

Ces données non structurées nouvellement nommées diffèrent :

- Images ;
- Courriels (avec analyse sémantique de leur contenu) ;
- Données issues des réseaux sociaux (commentaires et avis des internautes, par exemple sur Facebook ou Twitter) ;

Ces trois caractéristiques, décrites par trois "V", sont les principes qui définissent le Big Data. Il s'agit avant tout d'un changement d'orientation de l'utilisation des données. En bref, l'objectif principal du Big Data est de définir le Big Data, et il est donc nécessaire de l'analyser.

1.1.3 Architecture Big Data :

On distingue notamment les couches suivantes :

- Couche matérielle (couche infrastructure) : voir serveurs virtuel VMware ou serveurs lame;

-
- Couche de Stockage : les données sont stockées dans une base de données NoSQL, soit directement dans le système de fichiers distribué, soit dans le Datawarehouse ;
 - Couche de gestion et de traitement : dans cette couche, nous voyons des outils de traitement et d'analyse de données tels que MapReduce ou Pig.
 - Couche de visualisations : Pour la visualisation du résultat du traitement.

1.1.4 Avantages de l'architecture Big Data :

Il existe de nombreux avantages associés à l'architecture Big Data, tels que :

- Évolutivité : Quelle est la taille de votre infrastructure ? De combien d'espace disque avez-vous besoin maintenant et à l'avenir ? Le concept Big Data nous permet de surmonter ces problèmes car il fournit une architecture évolutive.
- Performances : Avec un traitement parallèle des données et un système de fichiers distribué, le concept Big Data est très efficace car il réduit la latence des applications.
- Faible coût : L'outil le plus important pour le Big Data, Hadoop, est Open Source, et il n'est pas nécessaire de centraliser les données dans un stockage souvent hors de prix grâce au Big Data et au système de fichiers distribué, disques de serveur interne.
- Disponibilité : Nous n'avons plus besoin de disques RAID, souvent coûteux.
- L'architecture Big Data possède ses propres mécanismes de haute disponibilité.

1.1.5 Domaines d'utilisation du Big Data :

Le Big Data a trouvé sa place dans de nombreux endroits :

Dans la première catégorie se trouvent les secteurs qui gèrent quotidiennement de gros volumes de données avec des problèmes de vitesse. Je serais là

- Banques: inviolabilité des anciennes données en raison des restrictions réglementaires;
- Télécommunications : analyse en temps réel de l'état du réseau;
- Médias numériques : ciblage publicitaire et analyse de sites Web;
- Marchés financiers : analyse des transactions pour la gestion des risques et des fraudes ainsi que l'analyse des clients.

Le secteur de la deuxième catégorie est hétérogène, les besoins mais aussi l'utilisation du Big Data peuvent fortement varier. On y trouvera :

- Services publics : analyse des compteurs (gaz, électricité...) et facilité management ;
- Marketing : ciblage publicitaire et analyse des tendances ;
- Santé : analyse du dossier médical et analyse génomique.

1.1.6 Les principales technologies de Big Data :

Il y a beaucoup d'entre eux. Diverses solutions peuvent être utilisées pour optimiser le temps de traitement des bases de données géantes :

- Les bases de données NoSQL (telles que MongoDB, Cassandra ou Redis) qui implémentent des systèmes de stockage considérés comme plus efficaces que SQL traditionnel pour l'analyse de données en masse (orienté clé/valeur, document, colonne ou graphique).
- Infrastructure serveur pour répartir le traitement sur des dizaines, des centaines, voire des milliers de nœuds. C'est ce qu'on appelle le traitement parallèle en masse.
- Le Framework Hadoop est probablement le plus connu d'entre eux. Combine le système de fichiers distribué HDFS, NoSQL HBase et l'algorithme MapReduce.
- Stockage des données : Nous parlons de traitement en mémoire, qui désigne un traitement effectué dans la mémoire d'un équipement informatique à accès aléatoire, au lieu de serveurs externes. L'avantage du traitement (en mémoire) est la rapidité car les données sont facilement accessibles. En revanche, ces données ne sont pas stockées longtemps, ce qui peut poser des problèmes d'historique.

1.1.7 Bases données NoSQL :

Les bases de données NoSQL (No-SQL ou Not Only SQL) sont actuellement un sujet brûlant. Le terme NoSQL fait référence à une catégorie de systèmes de gestion de bases de données conçus pour gérer de grandes bases de données pour plusieurs sites publics. Les bases de données NoSQL sont évolutives, elles peuvent traiter les données de manière distribuée. Les avantages de NoSQL incluent :

- Leurs performances ne s'effondrent jamais, quel que soit le degré de traitement. Leur temps de réaction est égal à un nombre ;
- Se déplace facilement. En effet, contrairement au SGBDR classique, il n'est pas nécessaire d'arrêter le service de déploiement de composants qui affecte les modèles de données ;
- Facile à mesurer. Par exemple, le plus grand cluster NoSQL a 400 To, tandis qu'Oracle peut traiter jusqu'à vingt téraoctets (pour un temps de réponse raisonnable).

1.1.7.1 Caractéristiques NoSQL

- Grande quantité de données ;
- Réplication et distribution évolutives ;
 - Des centaines de machines voire des milliers ;
 - Distribué dans le monde entier.
- Questions nécessitant une réponse rapide
- Asynchronisme des insertions et mises à jour ;
- Beaucoup de gens ne respectent pas l'acidité ;
- Performances en lecture/écriture ;
 - Des centaines de milliers de lectures par seconde ;
 - Des centaines de milliers d'écritures/secondes ;

1.1.7.2 Les types des bases NoSQL :

Il en existe 4 types distincts qui s'utilisent différemment et qui se prêtent mieux selon les types donnés que l'on souhaite y stocker.

- Clé-Valeur

Les BD NoSQL fonctionnant sur le principe Clé-Valeur sont les plus basiques que l'on peut trouver.

- Elles fonctionnent comme un grand tableau associatif et retourne une valeur dont elle ne connaît pas la structure ;
- Leur modèle peut être assimilé à une table de hachage (hashmap) distribuée ;
- Les données sont simplement représentées par un couple clé/valeur ;
- La valeur peut être une simple chaîne de caractères, ou un objet sérialisé.

- Document

Elles sont basées sur le modèle « clé-valeur » mais la valeur est un document en format semi-structuré hiérarchique de type JSON ou XML (possible aussi de stocker n'importe quel objet, via une sérialisation). Elles stockent une collection de "documents".

- Colonnes

Les données sont stockées dans des colonnes, pas dans des lignes, on peut facilement rajouter des colonnes dans les tableaux, par contre insérer une ligne revient plus cher, si les données dans la colonne sont les mêmes, on peut facilement compresser la colonne.

C'est le modèle près de la table dans RDBMS, mais voici le nombre de colonnes :

- est dynamique ;

- peut varier d'un enregistrement à l'autre pour éviter la recherche de colonnes NULL.

- Graphe

Ils permettent la modélisation, le stockage et la manipulation de données complexes associées à des relations insignifiantes ou variables.

- modèle de représentation des données basées sur la théorie des graphes

- basé sur les notions de nœuds, les relations et les propriétés qui leur sont associées.

1.1.7.3 Les principales bases de données NoSQL :

- MongoDB

La plus populaire des bases NoSQL documentaires est écrite en C et n'utilise pas de machine virtuelle JAVA. Cette base est idéale pour débiter car elle est à la fois polyvalente et simple. Aussi à l'aise pour le stockage massif de données que pour le développement rapide orienté Web. Elle possède également une documentation de premier ordre.

- Cassandra

Cassandra est le projet open source qui découle de la technologie de stockage Facebook. À l'origine, elle a été écrite spécifiquement pour répondre à la croissance explosive de cette entreprise. Elle est assez complexe à configurer, mais elle permet d'adresser toutes les situations où la performance et le traitement de la volumétrie est critique. Cassandra est une base de données en colonnes écrite en JAVA.

- HBase

Hbase est inspirée des publications de Google sur BigTable. Comme BigTable, elle est une base de données orientée colonne. Basée sur une architecture maître/esclave, les bases de données HBase sont capables de gérer d'énormes quantités d'informations (plusieurs milliards de lignes par table).

2. L'Annotation des images :

Définition : L'annotation peut être définie comme le processus de traduction explicite d'un document. La création de métadonnées à partir de documents annotés est l'une des techniques les plus importantes pour s'assurer que les données sont lisibles par machine.

Les métadonnées associées aux images peuvent être classées comme suit :

- Métadonnées de contenu indépendantes, où les métadonnées sont liées à l'image, telles que les noms d'auteur, la date, le lieu.

- Basé sur le contenu des métadonnées, où les métadonnées sont liées à des attributs de niveau bas et/ou moyen, par exemple : couleur, texture, forme.
- Métadonnées descriptives du contenu, où les métadonnées sont associées au contenu sémantique.

Le contenu des métadonnées descriptives peut être livré à deux niveaux spécifiques :

- Un contenu descriptif associé à l'ensemble de l'image.
- Partager une image avec des liens vers du contenu descriptif dans chaque section. Les systèmes d'annotation d'images à contenu descriptif basés sur des ontologies utilisent généralement deux types d'ontologies : l'une pour définir le schéma d'annotation et l'autre pour définir les concepts de domaine.

3.Introduction d'Hadoop :

Hadoop est le Framework Java open source d'Apache permettant de traiter de gros volumes de données de l'ordre de plusieurs pétaoctets (plusieurs milliers de To).

Hadoop a été conçu par Doug Cutting en 2004, qui a également créé le moteur Nutch Open Source. Doug Cutting cherche un moyen d'augmenter la taille d'index de sa machine. Il a eu l'idée de créer un cadre de gestion de fichiers distribués. Yahoo! après être devenu le principal contributeur, le portail a utilisé principalement l'infrastructure pour soutenir son moteur de recherche historique. Avec plus de 10 000 clusters Linux en 2008, c'était l'une des premières architectures Hadoop à mériter ce nom. Spécialement conçu pour les gros volumes. Facebook pour l'analyse des logs, Google pour l'analyse des besoins etc... Ceci est décrit comme suit :

- Stable : lorsqu'un nœud informatique redirige, ses tâches sont automatiquement distribuées aux autres nœuds. Les blocs de données sont également simulés ;
- Coûts : optimise les coûts en utilisant mieux les ressources présentées ;
- Flexible : parce qu'il répond à la nature des différentes données en traitant différents types de données ;
- Virtualisation : ne plus s'appuyer directement sur l'infrastructure physique (stockage coûteux), mais choisir de virtualiser ses clusters Hadoop. Il existe actuellement trois principales distributions Hadoop disponibles : Cloudera, Hortonworks, MapR.

Nous présenterons deux concepts de base de Hadoop : la version personnalisée de l'algorithme MapReduce, qui est Hadoop MapReduce, et le système de fichiers distribué HDFS.

3.1 Le système de fichier distribué d'Hadoop HDFS :

Hadoop utilise son propre système de fichiers virtuel : HDFS (Hadoop Distributed File System).

HDFS est un système de fichiers distribué, extensible et portable inspiré du système de fichiers Google (GFS).

Il est conçu pour stocker de très grandes quantités de données sur plusieurs machines équipées de disques durs classiques et permet l'abstraction de l'architecture de stockage physique pour manipuler un fichier système distribué comme s'il s'agissait d'un disque dur.

Cependant, HDFS diffère du système de fichiers classique pour les raisons suivantes :

- HDFS ne dépend pas du noyau du système d'exploitation. Il est pratique et peut être déployé sur une variété de systèmes d'exploitation. L'un de ses inconvénients est qu'il nécessite une application externe pour monter un lecteur de disque HDFS ;

-HDFS est un système de distribution dans un système conventionnel, la taille du disque est généralement considérée comme la limite d'utilisation globale. Dans HDFS, chaque nœud d'un cluster est un sous-ensemble de la quantité totale de données dans le cluster. Pour augmenter ce nombre global, il suffit d'ajouter de nouveaux nœuds. Nous voyons également HDFS, un service central appelé NameNode, qui gère les métadonnées.

- HDFS utilise une taille de bloc plus grande qu'un système conventionnel. La taille est généralement définie sur 64 Mo. Cependant, vous pouvez ajouter jusqu'à 128 Mo, 256 Mo, 512 Mo ou même 1 Go. Plus grand pour réduire le temps de blocage d'accès.

Notez que si la taille du fichier est inférieure à la taille du bloc, le fichier n'occupera pas toute la taille du bloc;

- HDFS fournit un système de réplication de blocs dont le nombre de répliques peut être configuré. Lors de la phase d'écriture, chaque bloc correspondant au fichier est répliqué sur plusieurs nœuds. Si un bloc n'est pas disponible sur un nœud pendant la phase de lecture, des copies de ce bloc sont disponibles sur les autres nœuds.

3.3.2 Les HDFS définit deux types de nœuds :

3.3.2.1 Le nœud principal ou NameNode

Il se caractérise par :

- Responsable de la distribution et de la réplication des blocs ;
- Serveur d'informations du Hdfs pour le client Hdfs ;
- Stocke et gère les métadonnées ;

- Comporte la liste des blocs pour chaque fichier (dans le cas de lecture) ;
- Contient la liste des DataNodes pour chaque bloc (dans le cas de l'écriture) ;
- Tenir les attributs des fichiers (ex : nom, date de création, facteur de réplication) ;
- Logs toute métadonnée et toute transaction sur un support persistant ;
- Lectures/écritures ;
- Créations/suppressions ;
- Démarre à partir d'une image d'HDFS (fsimage).

3.3.2.2 Le nœud de données ou DataNode

Il se caractérise par :

- Stocke des blocs de données dans le système de fichier local ;
- Maintenir des métadonnées sur les blocs possédés (ex : CRC) ;
- Serveur de bloc de données et de métadonnées pour le client hdfs ;
- Heartbeat avec le Namenode : Heartbeat est système permettant sous Linux la mise en clusters de plusieurs serveurs pour effectuer entre eux un processus de tolérance de panne. Le processus Heartbeat se chargera de passer un message-aller vers le NameNode indiquant : son identité, sa capacité totale, son espace utilisé, son espace restant.

3.3.2.3 Secondary NameNode

Le Name Node dans l'architecture Hadoop est un point unique de défaillance. Si ce service est arrêté, il n'y a pas moyen de pouvoir extraire les blocs d'un fichier donné. Pour résoudre ce problème, un NameNode secondaire appelé Secondary NameNode a été mis en place dans l'architecture Hadoop. Son rôle consiste à :

- Télécharger régulièrement les logs sur le NameNode ;
- Crée une nouvelle image en fusionnant les logs avec l'image HDFS ; composantes d'HDFS :
- Renvoie la nouvelle image au NameNode

3.3.3 Lecture d'un fichier HDFS :

Pour lire un fichier au sein de HDFS, il faut suivre les étapes suivantes :

Etape 1 : Le client indique au NameNode qu'il souhaite lire le fichier data.txt

Etape 2 : Le NameNode lui indiquera la taille de fichier (nombre de blocs) ainsi que les différents Data Node hébergeant les n blocs.

Etape 3 : Le client récupère chacun des blocs à un des DataNodes.

Etape 4 : En cas d'erreur/non réponse d'un des DataNode, il passe au suivant dans la liste fournie par le NameNode.

3.3.4 Ecriture dans un fichier HDFS

Etape 1 : On va utiliser la commande principale de gestion de Hadoop : Hadoop, avec l'option fs. Admettons qu'on souhaite stocker le fichier data.txt sur HDFS.

Etape 2 : Le programme va diviser le fichier en blocs de 64KB (ou autre, selon la configuration)

– supposons qu'on ait ici 3 blocs.

Etape 3 : Le NameNode lui indique les DataNodes à contacter.

Etape 4 : Le client contacte directement le DataNode concerné et lui demande de stocker le bloc.

Etape 5 : les DataNodes s'occuperont – en informant le NameNode – de répliquer les données entre eux pour éviter toute perte de données.

3.4 Ecosystèmes d'Hadoop :

Aujourd'hui il est difficile de se retrouver dans la " jungle " d'Hadoop pour les raisons suivantes :

- Ce sont des technologies jeunes ;
- Beaucoup de travaux d'acteurs différents (spécialistes du web, start-up,...) qui veulent prendre le train Big Data en marche ;

Dans une distribution Hadoop on va retrouver les éléments suivant sous leurs équivalences : HDFS, MapReduce, Zookeeper, Hbase, Hive, HCatalog, Oozie, Pig, Sqoop,...etc.

On peut toutefois distinguer trois distributions majeures qui sont Cloudera, Horton Works et MapR, toutes les trois se basant sur Apache Hadoop :

- MapR : Noyau Hadoop mais repackagé et enrichi de solutions propriétaires ;
- Cloudera : Fidèle et open source en grande partie sauf pour les outils d'administration ;
- Horton Works : Fidèle à la distribution Apache et donc 100% open source.

3.5 Mise en œuvre et l'installation de Hadoop :

Pour installer et créer un cluster Hadoop nous sommes besoins de :

- Java 8
- SSH
- PDSH

- Utilisation de trois machines virtuelles (VM) avec un système d'exploitation Linux Ubuntu 20.04

- Installation d'Hadoop à partir de la distribution Apache qui est considérée, actuellement, comme la distribution la plus utilisée dans les entreprises, du fait qu'elle propose une version open source complète qui utilise les principaux composants d'Hadoop (HDFS, MapReduce, Hbase, Pig, Zookeeper).

3.5.1 L'environnement de travail :

Dans la section suivante, on va décrire les outils et systèmes composant notre environnement de travail.

3.5.1.1 Virtual Box :

VirtualBox est un logiciel open source pour virtualiser l'architecture informatique x86-x64.

Il agit comme un hyperviseur, créant une VM (machine virtuelle) où l'utilisateur peut exécuter un autre système d'exploitation (Operating system).

3.5.1.2 Ubuntu:

Ubuntu : est un système d'exploitation¹⁾ libre, gratuit, sécurisé et convivial, qui peut facilement remplacer ou cohabiter avec un système actuel (Windows, MacOS, autre

GNU/Linux, ...). Avec Ubuntu, nous pouvons naviguer sur Internet, lire et écrire des courriels, créer des documents, des présentations et des feuilles de calculs, gérer votre bibliothèque multimédia et bien plus encore.

La distribution de base a eu trois périodes de diffusion : La période 1993-2004 : stabilisation du noyau Linux et apparition des variantes (Arch, Debian,...)

La période 2004-2017 : Debian sert de base à sa variante : Ubuntu. Unity est développée en vue de la convergence desktop-phone dans sa version 8.

La période 2017-actuelle : Unity est abandonnée au profit de Gnome.

Le terme « ubuntu » n'a pas été choisi au hasard : il véhicule des notions de communauté et de partage, qui sont les valeurs fondamentales du logiciel ouvert et libre.

Le mot « ubuntu » correspond à une idéologie provenant d'Afrique sub-saharienne qui s'articule autour des relations et des obligations des hommes les uns envers les autres.

Il recouvre un concept qui se développe depuis la Renaissance en Occident et qu'on appelle

Humanisme au XVI^{ème} siècle avec l'apparition de l'imprimerie. Cette technologie sera la pierre angulaire de la diffusion des connaissances pendant cette période dénommée « le Siècle des Lumières ». La philosophie qui caractérise cette époque est traduite dans la maxime : le salut est

dans le savoir et dans Autrui, il est donc indispensable d'acquérir la connaissance et de pratiquer l'ouverture d'esprit envers ses semblables.

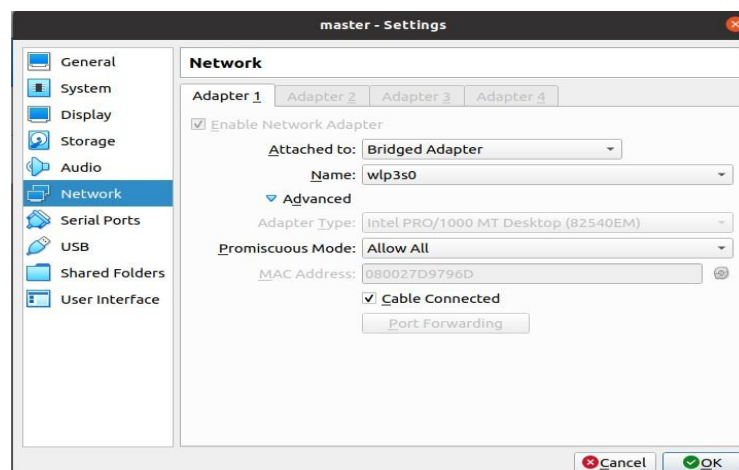
3.5.1.3 Choix d'Ubuntu 20.04:

Le système d'exploitation open source Ubuntu offre plusieurs opportunités, parmi lesquelles, On trouve:

- Facile à installer : un environnement graphique clair vous guide durant cette phase
- délicate et tout est prêt en 40 mn environ, y compris l'installation des dernières mises
- à jours de sécurité et corrections de paquets logiciels.
- Moins gourmand et plus léger : après installation du système et des applications,
- Ubuntu n'occupe que 6 Go sur mon disque dur, contre le double ou le triple pour un
- système tel que Windows 7. Ubuntu est également moins exigeant en termes de
- puissance et fonctionnera donc parfaitement sur un modèle d'ordinateur un peu ancien.
- est un système multi utilisateur.
- Pas de menaces et n'a pas besoin d'antivirus, il est sécurisé.
- est multitâche – gère plusieurs tâches en même temps comme Windows.
- Ubuntu vous propose de nombreux logiciels – Firefox, OpenOffice, logiciels de jeux,
- de formations, bureautiques, de sciences, d'éducatons etc...
- Il facilite la communication en ligne – Yahoo, Google, Skype.
- Gérer parfaitement comme tout système d'exploitation les images et vidéos.

3.5.1 Les étapes d'installation et configuration de l'environnement de travail

Etape 01



La configuration de notre réseau : où in est dit "Mode promiscuité", sélectionnez l'option "Autoriser tout".

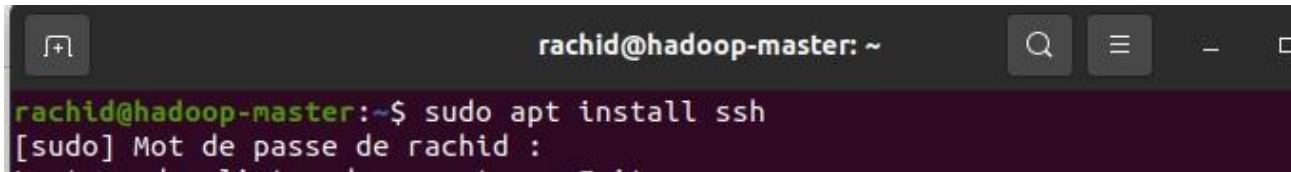
Etape 02.

Installez SSH à l'aide de la commande suivante :

```
sudo apt install ssh
```

Il vous demandera le mot de passe. Quand il demande une confirmation, donnez-le simplement.

Etape 03



```
rachid@hadoop-master: ~
rachid@hadoop-master:~$ sudo apt install ssh
[sudo] Mot de passe de rachid :
```

Installer PDSH a l'aide de la commande :

```
sudo apt install pdsh
```

Tout comme avant, donnez une conformation en cas de besoin

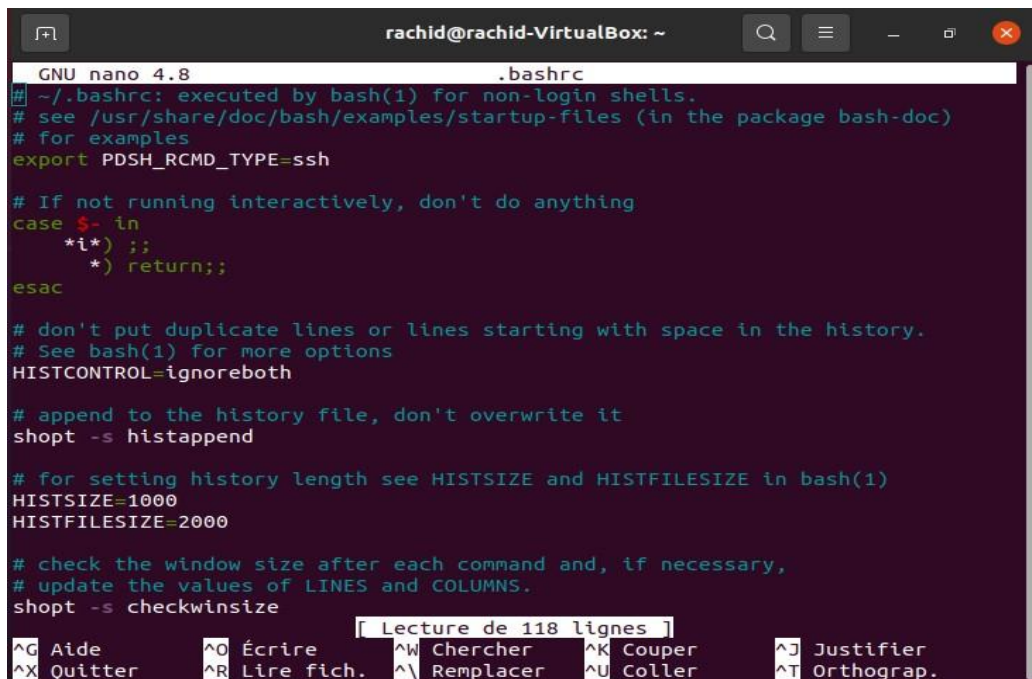
Etape 04

Ouvrez le fichier .bashrc avec la commande suivante :

```
nano .bashrc
```

A la fin du fichier, il suffit d'écrire la ligne suivante :

```
export PDSH_RCMD_TYPE=ssh
```



```
rachid@rachid-VirtualBox: ~
GNU nano 4.8 .bashrc
# ~/.bashrc: executed by bash(1) for non-login shells.
# see /usr/share/doc/bash/examples/startup-files (in the package bash-doc)
# for examples
export PDSH_RCMD_TYPE=ssh

# If not running interactively, don't do anything
case $- in
  *(i*) ;;
  *) return;;
esac

# don't put duplicate lines or lines starting with space in the history.
# See bash(1) for more options
HISTCONTROL=ignoreboth

# append to the history file, don't overwrite it
shopt -s histappend

# for setting history length see HISTSIZE and HISTFILESIZE in bash(1)
HISTSIZE=1000
HISTFILESIZE=2000

# check the window size after each command and, if necessary,
# update the values of LINES and COLUMNS.
shopt -s checkwinsize

Lecture de 118 lignes
^G Aide      ^O Écrire    ^W Chercher  ^K Couper    ^J Justifier
^X Quitter   ^R Lire fich.^_ Remplacer  ^U Coller    ^T Orthograp.
```

Etape 05

```

rachid@rachid-VirtualBox: ~
rachid@rachid-VirtualBox:~$ ssh-keygen -t rsa -P ""
Generating public/private rsa key pair.
Enter file in which to save the key (/home/rachid/.ssh/id_rsa):
Your identification has been saved in /home/rachid/.ssh/id_rsa
Your public key has been saved in /home/rachid/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:L80EXT4+3eyEBT/tsI/TFMW8BuVccZAl444179l+BvA rachid@rachid-VirtualBox
The key's randomart image is:
+---[RSA 3072]-----+
|                 *** |
|      .   oo= + |
|      o .  =.o. |
|      +   o=.o  |
|      . S  .=+oo |
|      + +  =E o  |
|      * +  =o.+  |
|      + o.o..o  |
|      o...o    |
+-----[SHA256]-----+
rachid@rachid-VirtualBox:~$ █

```

Configuration maintenant SSH, création d'une nouvelle clé à l'aide de la commande suivante :
Appuyez simplement sur Entrée à chaque fois que cela est nécessaire.

Etape 06

Maintenant, nous devons copier la clé publique dans le fichier `authorized_keys` avec la commande suivante :

```
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

```

|      + o.o..o  |
|      o...o    |
+-----[SHA256]-----+
rachid@rachid-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys █

```

Etape 07

Nous pouvons maintenant vérifier la configuration SSH en nous connectant à l'hôte local :

```
ssh localhost
```

Tapez simplement « yes » et appuyez sur Entrée si nécessaire.

```

rachid@rachid-VirtualBox:~$ cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
rachid@rachid-VirtualBox:~$ ssh localhost
The authenticity of host 'localhost (127.0.0.1)' can't be established.
ECDSA key fingerprint is SHA256:DMgT0X0cdWqFXJqZQJ4DLpa4/iDGJ4GhCcwBRCjTyik.
Are you sure you want to continue connecting (yes/no/[fingerprint])? █

```

Etape 08

C'est l'étape où nous installons Java 8. Nous utilisons cette commande

```
sudo apt install openjdk-8-jdk
```

Tout comme précédemment, donnez une confirmation en cas de besoin

```
Réception de :2 http://dz.archive.ubuntu.com/ubuntu focal-updates/universe amd64
4 openjdk-8-jre-headless amd64 8u292-b10-0ubuntu1~20.04 [28.2 MB]
9% [2 openjdk-8-jre-headless 4,263 kB/28.2 MB 15%] 322 kB/s 2min 1s
```

Etape 09

Cette étape n'est pas vraiment une étape, c'est juste pour vérifier si Java est maintenant correctement installé :

```
java -version
```

```
rachid@rachid-VirtualBox:~$ java -version
openjdk version "1.8.0_292"
OpenJDK Runtime Environment (build 1.8.0_292-8u292-b10-0ubuntu1~20.04-b10)
OpenJDK 64-Bit Server VM (build 25.292-b10, mixed mode)
rachid@rachid-VirtualBox:~$
```

Etape 10

Télécharger Hadoop à l'aide de la commande suivante :

```
sudo wget -P ~ https://mirrors.sonic.net/apache/hadoop/common/hadoop-3.3.1/hadoop-3.3
```

```
rachid@rachid-VirtualBox:~$ sudo wget -P ~ https://mirrors.sonic.net/apache/hado
op/common/hadoop-3.3.1/hadoop-3.3.1.tar.gz
[sudo] Mot de passe de rachid :
--2021-09-13 13:09:52-- https://mirrors.sonic.net/apache/hadoop/common/hadoop-3
.3.1/hadoop-3.3.1.tar.gz
Résolution de mirrors.sonic.net (mirrors.sonic.net)... 157.131.224.201, 2001:5a8
:601:4003:157:131:224:201
Connexion vers mirrors.sonic.net (mirrors.sonic.net)|157.131.224.201|:443... con
necté.
requête HTTP transmise, en attente de la réponse... 200 OK
Taille : 605187279 (577M) [application/x-gzip]
Enregistre : «/home/rachid/hadoop-3.3.1.tar.gz»

hadoop-3.3.1.tar.gz 35%[=====>] 202.65M 224KB/s tps 19m 12s
```

Etape 11

Nous devons décompresser le fichier hadoop-3.3.1.tar.gz avec la commande suivante :

```
tar xzf hadoop-3.3.1.tar.gz
```

```
rachid@rachid-VirtualBox:~$ tar xzf hadoop-3.3.1.tar.gz
rachid@rachid-VirtualBox:~$
```

Etape 12

Changez le nom du dossier hadoop-3.3.1 en hadoop (cela a rendu son utilisation plus facile

Utilisez cette commande :

```
mv hadoop-3.3.1 hadoop
```

```

rachid@rachid-VirtualBox:~$ mv hadoop-3.3.1 hadoop
rachid@rachid-VirtualBox:~$ █

```

Etape 13

Ouvrez le fichier `hadoop-env.sh` dans l'éditeur `nano` pour éditer `JAVA_HOME` :

`nano ~/hadoop/etc/hadoop/hadoop-env.sh` Collez cette ligne dans `JAVA_HOME`

`export JAVA_HOME=/usr/lib/jvm/java-8-openjdk-amd64/`

```

rachid@rachid-VirtualBox:~$ nano ~/hadoop/etc/hadoop/hadoop-env.sh
rachid@rachid-VirtualBox:~$ █

```

Etape 14

Remplacez le répertoire du dossier `hadoop` par `/usr/local/hadoop`. C'est la commande :

`sudo mv hadoop /usr/local/hadoop`

```

rachid@rachid-VirtualBox:~$ sudo mv hadoop /usr/local/hadoop
[sudo] Mot de passe de rachid :
rachid@rachid-VirtualBox:~$ █

```

Fournissez le mot de passe si nécessaire.

Etape 15

Ouvrez le fichier d'environnement sur `nano` avec cette commande :

`sudo nano /etc/environment`

```

rachid@rachid-VirtualBox:~$ sudo nano /etc/environment

```

Ensuite les configurations suivantes :

```

PATH="/usr/local/sbin:/usr/local/bin:/usr/sbin:/usr/bin:/sbin:/bin:/usr/games:/usr/local/games:/usr/local/hadoop/bin:/usr/local/hadoop/sbin"JAVA_HOME="/usr/lib/jvm/java-8-openjdk-amd64/jre"

```

Etape 16

Maintenant, nous allons ajouter un utilisateur appelé `hadoopuser`, et nous allons mettre ses configurations :

`sudo adduser hadoopuser`

Fournissez le mot de passe et vous pouvez laisser le reste vide, appuyez simplement sur Entrée.

```

rachid@rachid-VirtualBox:~$ sudo adduser hadoopuser
Ajout de l'utilisateur « hadoopuser » ...
Ajout du nouveau groupe « hadoopuser » (1001) ...
Ajout du nouvel utilisateur « hadoopuser » (1001) avec le groupe « hadoopuser »
...
Création du répertoire personnel « /home/hadoopuser »...
Copie des fichiers depuis « /etc/skel »...
Nouveau mot de passe :
Retapez le nouveau mot de passe :
passwd : le mot de passe a été mis à jour avec succès
Modification des informations relatives à l'utilisateur hadoopuser
Entrez la nouvelle valeur ou « Entrée » pour conserver la valeur proposée
Nom complet []:
N° de bureau []:
Téléphone professionnel []:
Téléphone personnel []:
Autre []:
Ces informations sont-elles correctes ? [0/n] 0
rachid@rachid-VirtualBox:~$

```

Tapez maintenant ces commandes :

```
sudo usermod -aG hadoopuser hadoopuser
```

```
sudo chown hadoopuser:root -R /usr/local/hadoop/
```

```
sudo chmod g+rwx -R /usr/local/hadoop/
```

```
sudo adduser hadoopuser sudo
```

```

rachid@rachid-VirtualBox:~$ sudo usermod -aG hadoopuser hadoopuser
rachid@rachid-VirtualBox:~$ sudo chown hadoopuser:root -R /usr/local/hadoop/
rachid@rachid-VirtualBox:~$ sudo chmod g+rwx -R /usr/local/hadoop/
rachid@rachid-VirtualBox:~$ sudo adduser hadoopuser sudo
Ajout de l'utilisateur « hadoopuser » au groupe « sudo »...
Ajout de l'utilisateur hadoopuser au groupe sudo
Fait.

```

Etape 17

Nous devons maintenant vérifier l'adresse IP de la machine :

```
ip addr
```

```

rachid@hadoop-master: ~
rachid@hadoop-master:~$ ip addr
1: lo: <LOOPBACK,UP,LOWER_UP> mtu 65536 qdisc noqueue state UNKNOWN group default qlen 1000
    link/loopback 00:00:00:00:00:00 brd 00:00:00:00:00:00
    inet 127.0.0.1/8 scope host lo
        valid_lft forever preferred_lft forever
    inet6 ::1/128 scope host
        valid_lft forever preferred_lft forever
Rhythmbox <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc fq_codel state UP group default qlen 1000
    link/ether 08:00:27:d9:79:6d brd ff:ff:ff:ff:ff:ff
    inet6 fe80::22fd:ccf:86e6:e2f1/64 scope link noprefixroute
        valid_lft forever preferred_lft forever
rachid@hadoop-master:~$

```

Maintenant, comme vous pouvez le voir, l'adresse IP est 192.168.43.238 ; n'oubliez pas que ce sera différent pour vous, vous devez agir en conséquence lorsque les adresses IP seront utilisées plus tard.

Notre réseau sera le suivant :

Master : 192.168.43.238

Slave1 : 192.168.43.239

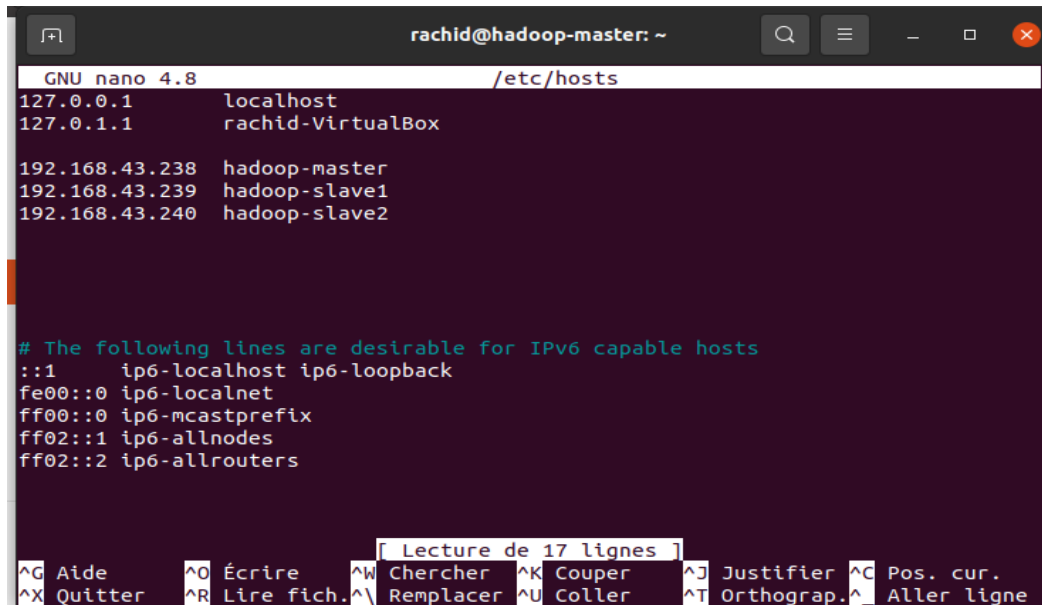
Slave2 : 192.168.43.240

Dans votre cas, continuez d'ajouter 1 au dernier numéro de l'adresse IP que vous obtenez sur votre machine, comme nous avons fait.

Etape 18

Ouvrez le fichier hosts et insérez vos configurations réseau :

```
sudo nano /etc/hosts
```



```
GNU nano 4.8 /etc/hosts
127.0.0.1 localhost
127.0.1.1 rachid-VirtualBox

192.168.43.238 hadoop-master
192.168.43.239 hadoop-slave1
192.168.43.240 hadoop-slave2

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

Lecture de 17 lignes
^G Aide      ^O Écrire    ^W Chercher  ^K Couper    ^J Justifier ^C Pos. cur.
^X Quitter   ^R Lire fich.^V Remplacer  ^U Coller    ^T Orthograp.^_ Aller ligne
```

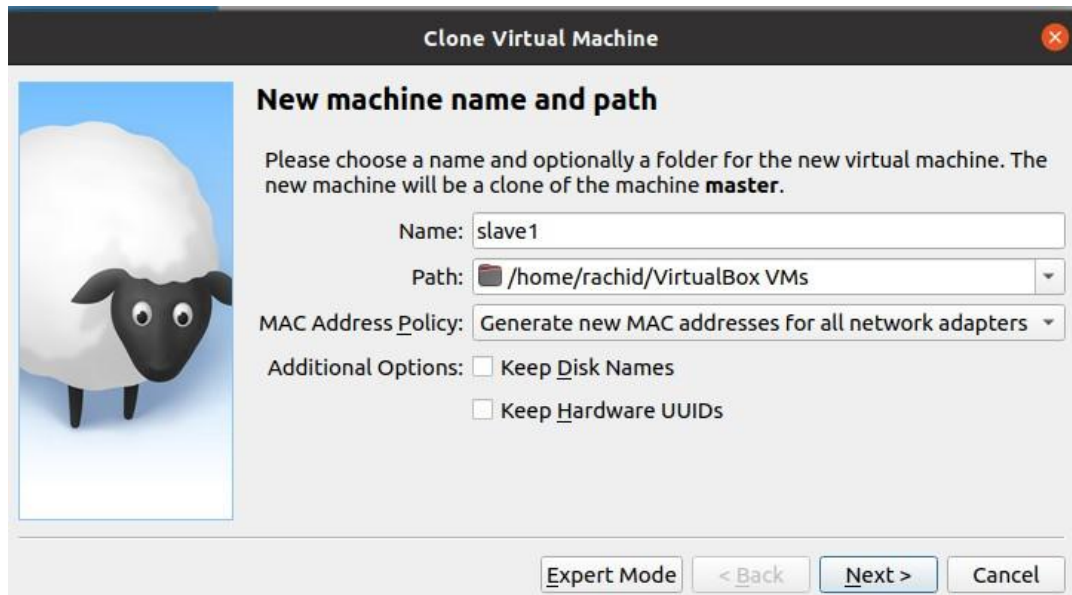
Etape 19

Il est maintenant temps de créer les masters.

Arrêtez votre machine virtuelle maître et clonez-la deux fois, en nommant un slave1 et l'autre slave2.

Assurez-vous que l'option « Générer de nouvelles adresses MAC pour toutes les cartes réseaux » est choisie.

Faites également un clone complet.



Etape 20

Sur la VM (machine virtuelle) principale, ouvrez le fichier de nom d'hôte sur nano :

```
sudo nano /etc/hostname
```

Insérez le nom de votre master machine virtuelle principale. (attention c'est le même nom que vous avez entrée précédemment dans le fichier hosts)

```

rachid@hadoop-master: ~
GNU nano 4.8 /etc/hosts
127.0.0.1 localhost
127.0.1.1 rachid-VirtualBox

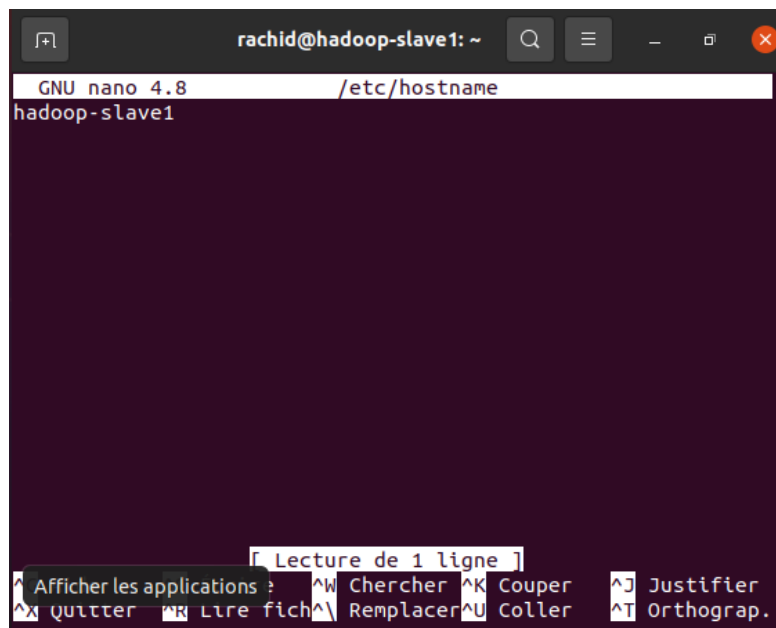
192.168.43.238 hadoop-master
192.168.43.239 hadoop-slave1
192.168.43.240 hadoop-slave2

# The following lines are desirable for IPv6 capable hosts
::1 ip6-localhost ip6-loopback
fe00::0 ip6-localnet
ff00::0 ip6-mcastprefix
ff02::1 ip6-allnodes
ff02::2 ip6-allrouters

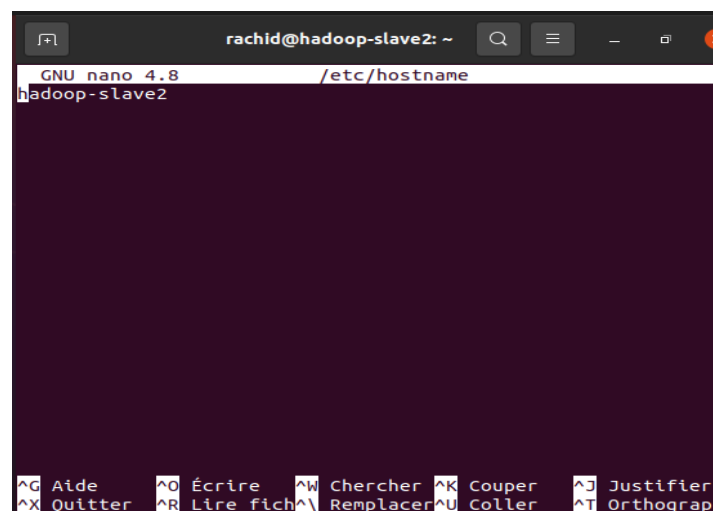
Lecture de 17 lignes
^G Aide ^O Écrire ^W Chercher ^K Couper ^J Justifier ^C Pos. cur.
^X Quitter ^R Lire fich. ^\ Remplacer ^U Coller ^T Orthograp. ^_ Aller ligne

```


Faires la même chose sur les slaves.



```
rachid@hadoop-slave1: ~  
GNU nano 4.8 /etc/hostname  
hadoop-slave1  
Lecture de 1 ligne  
Afficher les applications : Chercher Couper Justifier  
Quitter Lire fich Remplacer Coller Orthograp.
```



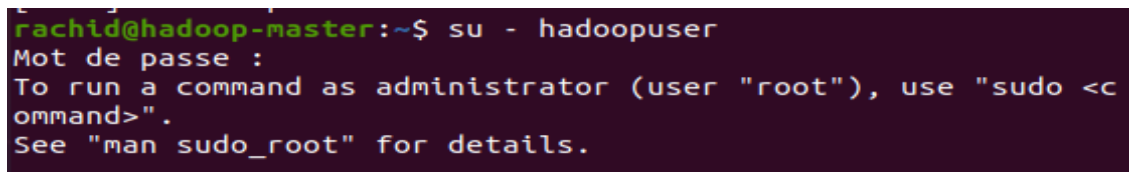
```
rachid@hadoop-slave2: ~  
GNU nano 4.8 /etc/hostname  
hadoop-slave2  
Aide Écrire Chercher Couper Justifier  
Quitter Lire fich Remplacer Coller Orthograp.
```

De plus, vous devez tous les redémarrer pour que cette configuration prenne effet : Sudo reboot

Etape21

Configurez le SSH sur hadoop-master, avec hadoopuser. C'est la commande :

sudo - hadoopuser



```
rachid@hadoop-master:~$ su - hadoopuser  
Mot de passe :  
To run a command as administrator (user "root"), use "sudo <command>".  
See "man sudo_root" for details.  
rachid@hadoop-master:~$ sudo - hadoopuser
```

Etape 22

Créer une clé SSH par la commande :

```
ssh-keygen -t rsa.
```

```
hadoopuser@hadoop-master:~$ ssh-keygen -t rsa
Generating public/private rsa key pair.
Enter file in which to save the key (/home/hadoopuser/.ssh/id_rsa):
Created directory '/home/hadoopuser/.ssh'.
Enter passphrase (empty for no passphrase):
Enter same passphrase again:
Your identification has been saved in /home/hadoopuser/.ssh/id_rsa
Your public key has been saved in /home/hadoopuser/.ssh/id_rsa.pub
The key fingerprint is:
SHA256:DZ3I1fdFXKNnjfBsxdvVPWPaBft0GYnw6hTVx50HL1o hadoopuser@hadoop-
master
The key's randomart image is:
+---[RSA 3072]----+
|                .o..oOO|
|      . + .+==@/|
|     + o. =XB&|
|      o ooE+*|
|     S .o o oo|
|      o . o |
|                . . |
+-----[SHA256]-----+
```

Etape 23

```
ssh-copy-id hadoopuser@hadoop-masterssh-copy-id hadoopuser@hadoop-slave1
ssh-copy-id hadoopuser@hadoop-slave2
```

Maintenant, nous devons copier la clé SSH pour tous les utilisateurs.

Utilisez cette commande :

```
hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-master
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/
hadoopuser/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-master (192.168.43.238)' can't be es
tablished.
ECDSA key fingerprint is SHA256:DMgT0X0cdWqFXJqZQJ4DLpa4/iDGJ4GhCcwBR
CjTyik.
Are you sure you want to continue connecting (yes/no/[fingerprint])?
yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s),
to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you
are prompted now it is to install the new keys
hadoopuser@hadoop-master's password:

Number of key(s) added: 1

Now try logging into the machine, with:  "ssh 'hadoopuser@hadoop-mas
ter'"
and check to make sure that only the key(s) you wanted were added.
```

```
hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-slave1
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/
hadoopuser/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-slave1 (192.168.43.239)' can't be es
tablished.
ECDSA key fingerprint is SHA256:DMgT0X0cdWqFXJqZQJ4DLpa4/iDGJ4GhCcwBR
CjTyik.
Are you sure you want to continue connecting (yes/no/[fingerprint])?
yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s),
to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you
are prompted now it is to install the new keys
hadoopuser@hadoop-slave1's password:

Number of key(s) added: 1

Now try logging into the machine, with:  "ssh 'hadoopuser@hadoop-sla
ve1'"
and check to make sure that only the key(s) you wanted were added.
```

```

hadoopuser@hadoop-master:~$ ssh-copy-id hadoopuser@hadoop-slave2
/usr/bin/ssh-copy-id: INFO: Source of key(s) to be installed: "/home/
hadoopuser/.ssh/id_rsa.pub"
The authenticity of host 'hadoop-slave2 (192.168.43.240)' can't be es
tablished.
ECDSA key fingerprint is SHA256:DMgT0X0cdWqFXJqZQJ4DLpa4/iDGJ4GhCcwBR
CjTyik.
Are you sure you want to continue connecting (yes/no/[fingerprint])?
yes
/usr/bin/ssh-copy-id: INFO: attempting to log in with the new key(s),
to filter out any that are already installed
/usr/bin/ssh-copy-id: INFO: 1 key(s) remain to be installed -- if you
are prompted now it is to install the new keys
hadoopuser@hadoop-slave2's password:

Number of key(s) added: 1

Now try logging into the machine, with:  "ssh 'hadoopuser@hadoop-sla
ve2'"
and check to make sure that only the key(s) you wanted were added.

```

Etape 24

Sur hadoop-master, ouvrez le fichier core-site.xml sur nano :

```
sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml
```

```

hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/core-site.xml
[sudo] password for hadoopuser:

```

Ajoutez ensuite les configurations suivantes :

```

<configuration>
  <property>
    <name>fs.defaultFS</name>
    <value>hdfs://hadoop-master:9000</value>
  </property>
</configuration>.

```

```
Fichier Édition Affichage Rechercher Terminal Aide
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/core-site.xml
Licensed under the Apache License, Version 2.0 (the "License");
you may not use this file except in compliance with the License.
You may obtain a copy of the License at

    http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or im
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
-->

<!-- Put site-specific property overrides in this file. -->

<configuration>
<property>
<name>fs.defaultFS</name>
<value>hdfs://hadoop-master:9000</value>
</property>
</configuration>

```

Etape 25

Toujours sur hadoop-master, ouvrez le fichier hdfs-site.xml :

```
sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
```

```
hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/hdfs-site.xml
1
```

Ajoutez les configurations suivantes:

```
<configuration>
<property>
<name>dfs.namenode.name.dir</name><value>/usr/local/hadoop/data/nameNode</value>
</property>
<property>
<name>dfs.datanode.data.dir</name><value>/usr/local/hadoop/data/dataNode</value>
</property>
<property>
<name>dfs.replication</name>
<value>2</value>
</property>
</configuration>
```

Chapitre I Big Data et Annotation Vers Hadoop

```
Fichier Édition Affichage Rechercher Terminal Aide
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/hdfs-site.xml Modifié
http://www.apache.org/licenses/LICENSE-2.0

Unless required by applicable law or agreed to in writing, software
distributed under the License is distributed on an "AS IS" BASIS,
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or im>
See the License for the specific language governing permissions and
limitations under the License. See accompanying LICENSE file.
->

!-- Put site-specific property overrides in this file. -->

configuration>
property>
name>dfs.namenode.name.dir</name><value>/usr/local/hadoop/data/name>
/property>
property>
name>dfs.datanode.data.dir</name><value>/usr/local/hadoop/data/data>
/property>
property>
name>dfs.replication</name>
value>2</value>
/property>

G Aide      ^O Écrire    ^W Chercher  ^K Couper    ^J Justifier
X Quitter   ^R Lire fich.^_ Remplacer ^U Coller   ^T Orthograp.
```

Etape 26

On est toujours sur hadoop-master, ouvrez le fichier **workers** :

```
sudo nano /usr/local/hadoop/etc/hadoop/workers
```

```
hadoopuser@hadoop-master:~$ sudo nano /usr/local/hadoop/etc/hadoop/workers
```

Ajoutez ces deux lignes: (les noms des slaves, vous vous souvenez du fichier hosts ?)

```
hadoop-slave1
```

```
hadoop-slave2
```

```
Fichier Édition Affichage Rechercher Terminal Aide
GNU nano 4.8 /usr/local/hadoop/etc/hadoop/workers
localhost
hadoop-slave1
hadoop-slave2

[ Lecture de 3 lignes ]

^G Aide      ^O Écrire    ^W Chercher  ^K Couper    ^J Justifier
^X Quitter   ^R Lire fich.^_ Remplacer ^U Coller   ^T Orthograp.
```

Etape 27

Vous devez copier les configurations hadoop-master sur les slaves pour cela vous utilisez ces commandes :

```
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave1:/usr/local/hadoop/etc/hadoop/
scp /usr/local/hadoop/etc/hadoop/* hadoop-slave2:/usr/local/hadoop/etc/hadoop/
```

Chapitre I Big Data et Annotation Vers Hadoop

```
rachid@hadoop-master: ~  
7061 DataNode  
6935 NameNode  
10778 Jps  
9117 NodeManager  
rachid@hadoop-master:~$ export HADOOP_HOME="/usr/local/hadoop"  
rachid@hadoop-master:~$ export HADOOP_COMMON_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop  
rachid@hadoop-master:~$ export HADOOP_HDFS_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_MAPRED_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_YARN_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ start-yarn.sh  
Starting resourcemanager  
Starting nodemanagers  
hadoop-slave2: nodemanager is running as process 4043. Stop it first  
and ensure /tmp/hadoop-rachid-nodemanager.pid file is empty before r  
etry.  
pdsh@hadoop-master: hadoop-slave2: ssh exited with exit code 1  
hadoop-slave1: nodemanager is running as process 3987. Stop it first  
and ensure /tmp/hadoop-rachid-nodemanager.pid file is empty before r  
etry.  
pdsh@hadoop-master: hadoop-slave1: ssh exited with exit code 1  
localhost: nodemanager is running as process 9117. Stop it first and  
ensure /tmp/hadoop-rachid-nodemanager.pid file is empty before retry  
.  
pdsh@hadoop-master: localhost: ssh exited with exit code 1  
rachid@hadoop-master:~$
```

Etape 28

Vous devez maintenant formater le système de fichiers HDFS, Exécutez ces commandes :

```
source /etc/environment  
hdfs namenode -format
```

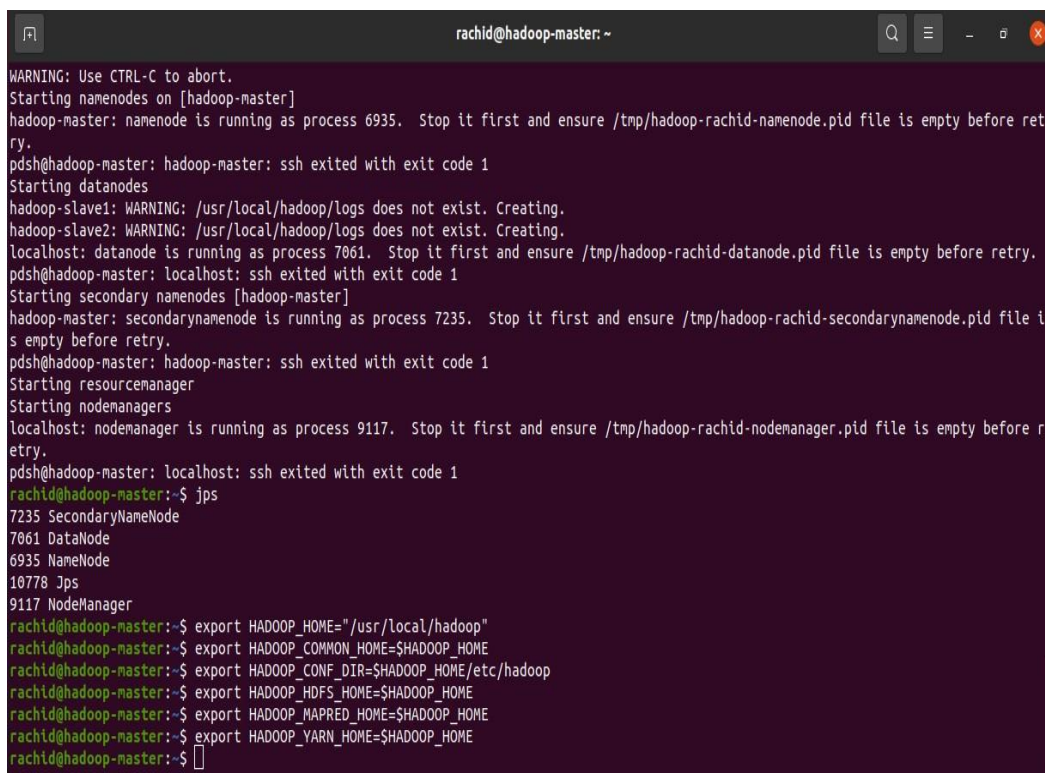
```
hadoopuser@hadoop-master: ~  
Fichier  Édition  Affichage  Rechercher  Terminal  Aide  
hadoopuser@hadoop-master:~$ hdfs namenode -format  
WARNING: /usr/local/hadoop/logs does not exist. Creating.  
2021-09-14 13:03:37,396 INFO namenode.NameNode: STARTUP_MSG:  
/*****  
STARTUP_MSG: Starting NameNode  
STARTUP_MSG:   host = hadoop-master/192.168.43.238  
STARTUP_MSG:   args = [-format]  
STARTUP_MSG:   version = 3.3.1  
STARTUP_MSG:   classpath = /usr/local/hadoop/etc/hadoop:/usr/local/ha  
doop/share/hadoop/common/lib/netty-3.10.6.Final.jar:/usr/local/hadoop  
/share/hadoop/common/lib/snappy-java-1.1.8.2.jar:/usr/local/hadoop/sh  
are/hadoop/common/lib/commons-lang3-3.7.jar:/usr/local/hadoop/share/h  
adoop/common/lib/jetty-io-9.4.40.v20210413.jar:/usr/local/hadoop/shar  
e/hadoop/common/lib/re2j-1.1.jar:/usr/local/hadoop/share/hadoop/commo  
n/lib/kerby-asn1-1.0.1.jar:/usr/local/hadoop/share/hadoop/common/lib/  
commons-beanutils-1.9.4.jar:/usr/local/hadoop/share/hadoop/common/lib  
/jackson-core-2.10.5.jar:/usr/local/hadoop/share/hadoop/common/lib/je  
tty-util-9.4.40.v20210413.jar:/usr/local/hadoop/share/hadoop/common/l  
ib/jackson-jaxrs-1.9.13.jar:/usr/local/hadoop/share/hadoop/common/lib  
/commons-codec-1.11.jar:/usr/local/hadoop/share/hadoop/common/lib/je  
rsey-servlet-1.19.jar:/usr/local/hadoop/share/hadoop/common/lib/zookee  
per-jute-3.5.6.jar:/usr/local/hadoop/share/hadoop/common/lib/guava-27  
.0-jre.jar:/usr/local/hadoop/share/hadoop/common/lib/asm-5.0.4.jar:/u  
sr/local/hadoop/share/hadoop/common/lib/dnsjava-2.1.7.jar:/usr/local/  
hadoop/share/hadoop/common/lib/hadoop-shaded-guava-1.1.1.jar:/usr/loc  
al/hadoop/share/hadoop/common/lib/commons-logging-1.1.3.jar:/usr/loca
```


Chapitre I Big Data et Annotation Vers Hadoop

Etape 29

Démarrez HDFS avec cette commande:

```
start-dfs.sh
```



```
rachid@hadoop-master: ~  
WARNING: Use CTRL-C to abort.  
Starting namenodes on [hadoop-master]  
hadoop-master: namenode is running as process 6935. Stop it first and ensure /tmp/hadoop-rachid-namenode.pid file is empty before re  
try.  
pdsh@hadoop-master: hadoop-master: ssh exited with exit code 1  
Starting datanodes  
hadoop-slave1: WARNING: /usr/local/hadoop/logs does not exist. Creating.  
hadoop-slave2: WARNING: /usr/local/hadoop/logs does not exist. Creating.  
localhost: datanode is running as process 7061. Stop it first and ensure /tmp/hadoop-rachid-datanode.pid file is empty before re  
try.  
pdsh@hadoop-master: localhost: ssh exited with exit code 1  
Starting secondary namenodes [hadoop-master]  
hadoop-master: secondarynamenode is running as process 7235. Stop it first and ensure /tmp/hadoop-rachid-secondarynamenode.pid file i  
s empty before retry.  
pdsh@hadoop-master: hadoop-master: ssh exited with exit code 1  
Starting resourcemanager  
Starting nodemanagers  
localhost: nodemanager is running as process 9117. Stop it first and ensure /tmp/hadoop-rachid-nodemanager.pid file is empty before r  
etry.  
pdsh@hadoop-master: localhost: ssh exited with exit code 1  
rachid@hadoop-master:~$ jps  
7235 SecondaryNameNode  
7061 DataNode  
6935 NameNode  
10778 Jps  
9117 NodeManager  
rachid@hadoop-master:~$ export HADOOP_HOME="/usr/local/hadoop"  
rachid@hadoop-master:~$ export HADOOP_COMMON_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop  
rachid@hadoop-master:~$ export HADOOP_HDFS_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_MAPRED_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_YARN_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$
```

Pour vérifier si cela a fonctionné, exécutez la commande suivante : cela vous indiquera quelles ressources ont été initialisées :

```
jps
```

Maintenant, vous devez faire la même chose dans les slaves :



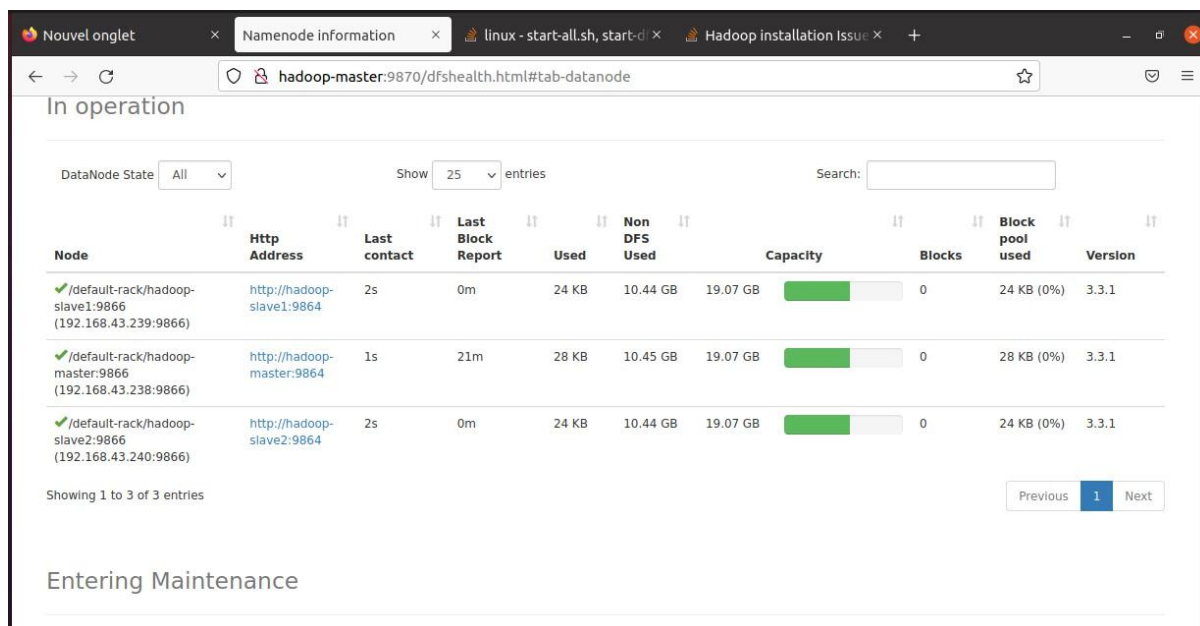
```
rachid@hadoop-slave1:~$ jps  
3987 NodeManager  
3845 DataNode  
4071 Jps  
rachid@hadoop-slave1:~$  
  
rachid@hadoop-slave2:~$ jps  
4043 NodeManager  
3901 DataNode  
4127 Jps  
rachid@hadoop-slave2:~$
```


Chapitre I Big Data et Annotation Vers Hadoop

Etape 30

Voyez si cela a fonctionné :

Ouvrez votre navigateur et tapez : `hadoop-master :9870`. C'est ce que vous verrez.



Comme vous pouvez le voir, les deux nœuds sont opérationnels !

Etape 31

Configuration **Yarn**, exécutez simplement les commandes suivantes :

```
export HADOOP_HOME="/usr/local/hadoop"  
export HADOOP_COMMON_HOME=$HADOOP_HOME  
export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop  
export HADOOP_HDFS_HOME=$HADOOP_HOME  
export HADOOP_MAPRED_HOME=$HADOOP_HOME
```

```
export HADOOP_YARN_HOME=$HADOOP_HOME
```

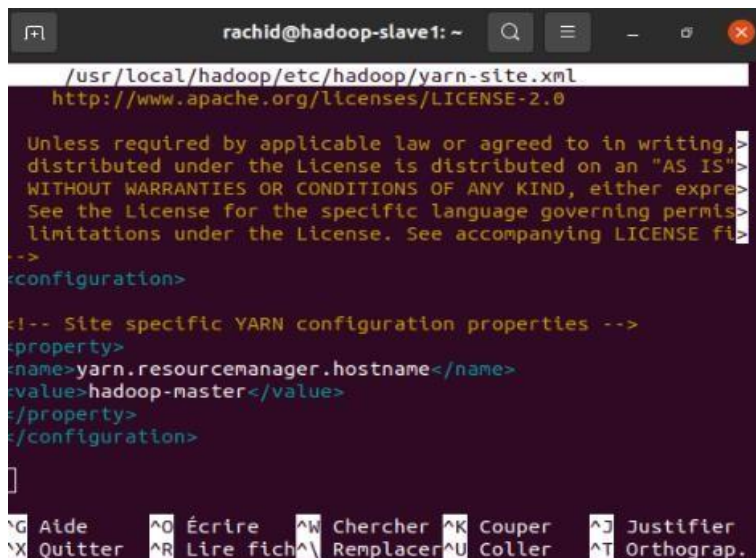
```
rachid@hadoop-master:~$ export HADOOP_HOME="/usr/local/hadoop"  
rachid@hadoop-master:~$ export HADOOP_COMMON_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_CONF_DIR=$HADOOP_HOME/etc/hadoop  
rachid@hadoop-master:~$ export HADOOP_HDFS_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_MAPRED_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$ export HADOOP_YARN_HOME=$HADOOP_HOME  
rachid@hadoop-master:~$
```

Etape 32

Dans les deux slaves, ouvrez **fil-site.xml** sur nano :

```
<property>  
<name>yarn.resourcemanager.hostname</name>  
<value>hadoop-master</value>  
</property>
```

Chapitre I Big Data et Annotation Vers Hadoop



```
rachid@hadoop-slave1: ~  
/usr/local/hadoop/etc/hadoop/yarn-site.xml  
http://www.apache.org/licenses/LICENSE-2.0  
Unless required by applicable law or agreed to in writing,  
distributed under the License is distributed on an "AS IS"  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either ex  
See the License for the specific language governing permis  
limitations under the License. See accompanying LICENSE fi  
-->  
<configuration>  
!-- Site specific YARN configuration properties -->  
<property>  
<name>yarn.resourcemanager.hostname</name>  
<value>hadoop-master</value>  
</property>  
</configuration>
```

sudo nano /usr/local/hadoop/etc/hadoop/yarn-site.xml Vous devez ajouter les configurations suivantes sur les deux slaves:



```
rachid@hadoop-slave2: ~  
/usr/local/hadoop/etc/hadoop/yarn-site.xml  
http://www.apache.org/licenses/LICENSE-2.0  
Unless required by applicable law or agreed to in writing,  
distributed under the License is distributed on an "AS IS"  
WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either ex  
See the License for the specific language governing permis  
limitations under the License. See accompanying LICENSE fi  
-->  
<configuration>  
<property>  
<name>yarn.resourcemanager.hostname</name>  
<value>hadoop-master</value>  
</property>  
!-- Site specific YARN configuration properties -->  
</configuration>
```

Etape 33

Sur le master, commencez-vous le fil. Utilisez cette commande :

start-yarn.sh

```
rachid@hadoop-master:~$ start-yarn.sh  
Starting resourcemanager  
Starting nodemanagers
```

Chapitre I Big Data et Annotation Vers Hadoop

Etape 34

Ouvrez votre navigateur. Vous allez maintenant taper : <http://hadoop-master:8088/cluster>.

The screenshot shows the Hadoop cluster management interface. The browser address bar displays `hadoop-master:8088/cluster/nodes`. The page title is "Nodes of the cluster". On the left, there is a sidebar with navigation options like "Cluster", "About", "Nodes", "Node Labels", "Applications", "NEW SAVINGS", "SUBMITTED", "ACCEPTED", "RUNNING", "FINISHED", "FAILED", "KILLED", "Scheduler", and "Tools". The main content area displays various metrics and a table of active nodes.

Cluster Metrics

Apps Submitted	Apps Pending	Apps Running	Apps Completed	Containers Running	Used Resources	Total Resources	Reserved Resources	Physical Mem Used %	Physical VCores Used %
0	0	0	0	0	<memory:0 B, vCores:0>	<memory:24 GB, vCores:24>	<memory:0 B, vCores:0>	77	8

Cluster Nodes Metrics

Active Nodes	Decommissioning Nodes	Decommissioned Nodes	Lost Nodes	Unhealthy Nodes	Rebooted Nodes	Shutdown Nodes
3	0	0	0	0	0	0

Scheduler Metrics

Scheduler Type	Scheduling Resource Type	Minimum Allocation	Maximum Allocation	Maximum Cluster Application Priority	Scheduler Busy %
Capacity Scheduler	[memory-mb (unit=M), vcores]	<memory:1024, vCores:1>	<memory:8192, vCores:4>	0	0

Nodes Table

Node Labels	Rack	Node State	Node Address	Node HTTP Address	Last health-update	Health-report	Containers	Allocation Tags	Mem Used	Mem Avail	Phys Mem Used %	VCores Used	VCores Avail	Phys VCores Used %	Version
/default-rack	/default-rack	RUNNING	hadoop-slave1:37745	hadoop-slave1:8042	mar. sept. 14 16:51:43 +0100 2021		0		0 B	8 GB	69	0	8	0	3.3.1
/default-rack	/default-rack	RUNNING	hadoop-master:33221	hadoop-master:8042	mar. sept. 14 16:52:03 +0100 2021		0		0 B	8 GB	82	0	8	29	3.3.1
/default-rack	/default-rack	RUNNING	hadoop-slave2:33021	hadoop-slave2:8042	mar. sept. 14 16:51:42 +0100 2021		0		0 B	8 GB	80	0	8	0	3.3.1

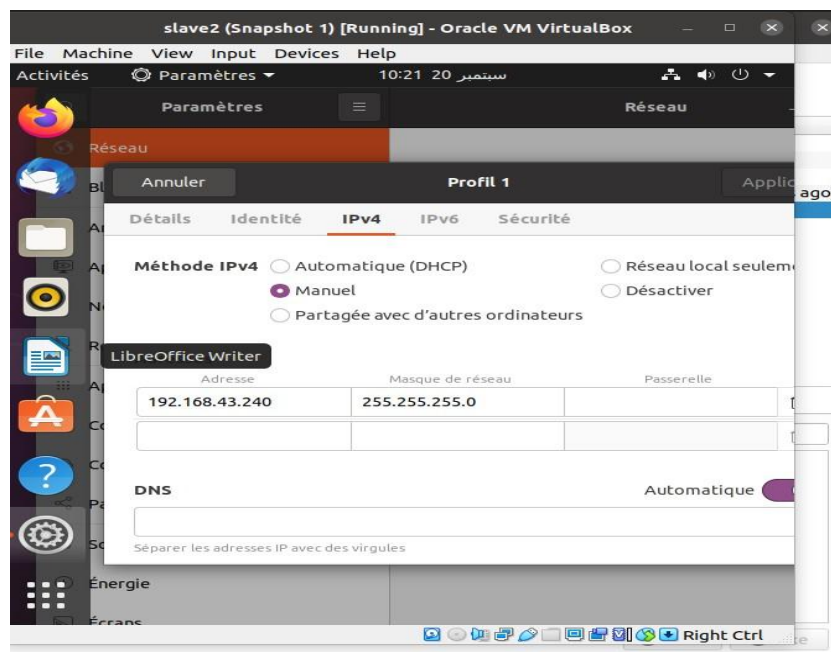
Comme vous pouvez le voir, le cluster affiche 3 nœuds actifs !

Les problèmes rencontrés :

The screenshot shows a Linux desktop environment with a network configuration window open. The window title is "slave1 (Snapshot 1) [Running] - Oracle VM VirtualBox". The window is titled "Paramètres" and shows the "Réseau" (Network) settings for "Profil 1". The "Méthode IPv4" (IPv4 Method) is set to "Manuel" (Manual). The "Adresses" (Addresses) section shows the IP address "192.168.43.239" and the netmask "255.255.255.0". The "DNS" section is set to "Automatique" (Automatic).

Dans l'étape 23, si vous ne pouvez pas copier la clé SSH pour tous les utilisateurs, vous avez comme solution de vérifier les adresses IP sur les slaves.

Chapitre I Big Data et Annotation Vers Hadoop



Dans l'étape 29 un problème de fonctionnement de démarrage de HDFS, vous pouvez le résoudre en définissant les autorisations de tous les fichiers sur 777 :

Utilisez cette commande :

```
sudo chmod 777 /usr/local/hadoop-3.3.1/* -R.
```

CHAPITRE II

Etude de cas Le

Cancer dans

l'Algérie

Introduction :

Depuis son existence, l'être humain est exposé à des pathologies diverses, certaines disparaissent avec le temps, pendant que d'autres résistent et prennent de l'ampleur dans l'environnement médico-social et médico-sanitaire, touchant ainsi toutes les couches sociales, mais à des degrés différents. Les répercussions de ces pathologies sur notre mode de vie au quotidien et l'augmentation de l'incidence de certaines maladies sont devenues une préoccupation majeure de santé publique. Cette tendance ascendante et négative dans le champ médical et le paysage sanitaire, a fait naître une certaine conscience humaine, personnelle, publique et politique ; pour une prise en charge précoce et adéquate, à travers notamment l'octroi de moyens financiers, humains et technologiques considérables, dans une stratégie nationale dans la prévention, la détection et la thérapie.

Le cancer est un problème majeur de santé publique dans le monde. Selon l'OMS, le nombre de nouveaux cas de cancer en 2012 est estimé à 14,1 millions et le nombre de décès à 8,2 millions.

Il a été estimé en 2008 que 70% des décès par cancer dans le monde survenaient dans les pays en développement. La fréquence des cancers pourrait augmenter de 50 % dans le monde, avec 15 millions de nouveaux cas par an en 2020.

L'Algérie a connu à partir des années 90, une transition démographique profonde et rapide qui a entraîné une modification structurelle du pro-épidémiologique de sa population. Celle-ci a connu une baisse de la mortalité générale qui a été divisée par 4 en l'espace de 50 ans (16,45 pour mille à la fin des années 60 à 4,41 pour mille habitants en 2008) et une baisse importante de la mortalité infanto-juvénile corrélée à une augmentation progressive de l'espérance de vie estimée à 25 années au cours des 50 dernières années, ce qui a eu pour conséquence un vieillissement progressif de la population avec une part de plus en plus importante des personnes âgées de plus de 60 ans dans la pyramide des âges.

Par ailleurs, une modification profonde du mode de vie collectif et individuel (augmentation du tabagisme, du stress, de la sédentarité, de l'urbanisation...) et d'un mode alimentaire déséquilibré sont à l'origine de l'émergence des Maladies Non Transmissibles (MNT) dont le cancer. Ces maladies constituent aujourd'hui plus de 80% des causes de maladies et ont en commun un certain nombre de facteurs de risque d'où la nécessité d'une politique commune de prévention contre ceux-ci.

2. Généralité sur le cancer :

2.1. Définition

Le cancer n'est pas une affection récente. Ce sont les prêtres d'Esculape et Hippocrate qui lui ont donné son nom : carcinons, pour « crabe » ou « pince », en raison de son aspect lorsqu'il est très évolué, les veines qui immergent autour de la tumeur étant alors comparables aux pattes d'un crabe [28].

Le cancer correspond à la prolifération cellulaire anarchique non contrôlée de cellules dites malignes, en raison de certains caractères anormaux. Commence comme une seule cellule qui perd la régulation du contrôle de la croissance, elle se divise pour faire de nombreuses cellules filles qui se développent en une tumeur. La prolifération peut rester localisée (tumeur) ou se propager dans d'autres sites, soit par voie lymphatique (envahissement ganglionnaire), soit par voie sanguine (métastases) [27].

La tumeur peut être :

- ✓ Tumeur maligne ou cancéreuse lorsque les cellules, en se développant, envahissent les organes avoisinants et peuvent envoyer par voie lymphatique ou sanguine des cellules dans des organes à distance. Ces colonies de cellules pourront se développer à leur tour et former ainsi des métastases [29].

- ✓ Tumeur bénigne où le tissu tumoral reproduit de très près la structure du tissu initial.

Dans ce dernier, les cellules ont une morphologie normale et ne présentent aucun caractère de malignité, et donc il n'y a pas d'envahissement des tissus voisins [30].

L'extension à distance, de façon discontinue par rapport à la tumeur originale, aboutissant à la formation de nouveaux foyers tumoraux appelés métastases, est une caractéristique majeure des cancers et la principale cause de leur malignité. Si les mécanismes de la prolifération cellulaire aboutissant à une tumeur sont maintenant bien compris sur le plan moléculaire, les mécanismes conduisant de la tumeur primitive à la métastase commencent seulement à être appréhendés correctement. Les cellules cancéreuses empruntent deux voies principales de dissémination :

- ✓ la voie lymphatique, conduisant à l'invasion des ganglions lymphatiques qui drainent la région où évolue la tumeur ;
- ✓ la voie sanguine, conduisant à l'invasion d'organes à distance comme le foie, le cerveau, l'os ou le poumon [31].

D'après **Couso**, [32] la formation de métastases résulte d'un processus complexe se déroulant en plusieurs étapes successives : les cellules tumorales acquièrent la

capacité de se détacher de la tumeur primitive et d'envahir le tissu adjacent ; elles atteignent la circulation et échappent à la surveillance immunitaire ; puis elles vont se nicher au niveau d'un organe à distance au niveau duquel elles vont pouvoir rester dormantes ou ré-initier leur prolifération pour former des foyers secondaires.

2.2. Epidémiologie du cancer

Plus de 60% des nouveaux cas de cancer surviennent en Afrique, en Asie, en Amérique centrale et en Amérique latine. Ces régions représentent 70% des décès par cancer dans le monde [32]. En Algérie, chaque année on estime le nombre de nouveaux cas de cancer à 50 000. Ce chiffre ne cesse d'augmenter chaque année du fait de la transition épidémiologique [33].

Cette augmentation du nombre de cas de cancers serait principalement due, à l'importance du tabagisme ainsi qu'à l'adoption croissante de modes de vie préjudiciables à la santé [34]. Ainsi, bien que la révolution industrielle et le développement de la médecine aient permis l'amélioration des conditions de vie des populations en diminuant la mortalité par les maladies transmissibles, ils ont contribué à l'allongement de l'espérance de vie avec apparition de nouveaux problèmes de santé publique et de maladies non transmissibles entre autres les cancers [35].

2.3. Les facteurs de risque de cancer

Le tabagisme, la consommation d'alcool, une mauvaise alimentation, un manque d'activité physique et la pollution de l'air sont autant de facteurs de risque de cancer (et d'autres maladies non transmissibles).

Certaines infections chroniques constituent elles aussi des facteurs de risque de cancer, en particulier dans les pays à revenu faible ou intermédiaire. Environ 13 % des cancers diagnostiqués dans le monde en 2018 étaient imputables à des infections cancérogènes, notamment celles dues à *Helicobacter pylori*, au papillomavirus humain (PVH), au virus de l'hépatite B, au virus de l'hépatite C et au virus d'Epstein Barr (2).

Les virus de l'hépatite B et de l'hépatite C augmentent le risque de cancer du foie, tandis que certains types de PVH majorent le risque de cancer du col de l'utérus. L'infection à VIH multiplie par six le risque de développer un cancer du col de l'utérus et accroît fortement le risque de développer certains autres cancers comme le sarcome de Kaposi.

2.4. Cancérogenèse

La cancérogenèse se réalise en plusieurs étapes dont certaines sont irréversibles, et d'autres réversibles [36].

Tout d'abord, l'initiation. C'est la première phase. Elle ne concerne qu'une seule cellule. Elle rend la cellule immortelle. On suppose que cette phase n'est due qu'à un seul facteur (chimique, physique, ou génétique) et que ce phénomène ne survient qu'une seule fois. Puis, survient l'étape de promotion : Au cours de la deuxième phase, la cellule acquiert par mutations successives, les caractéristiques qui lui permettent de créer un cancer. Ces étapes peuvent être réversibles, et sont modulées par des nombreux facteurs immunitaires et hormonaux.

Enfin, l'étape de progression cellulaire : Pendant cette phase, les cellules filles de la cellule transformée sont sélectionnées pour donner des clones plus malins et pour acquérir des propriétés leur permettant par exemple de métastaser [37].

Ayant acquis la capacité d'échapper à tout contrôle tissulaire, la cellule mutée débute sa prolifération. Celle-ci commence par une déstructuration du tissu environnant. Des protéases sont alors exprimées afin de digérer les protéines responsables de l'adhérence intercellulaire et de permettre ainsi de dissocier les épithéliums. Ces enzymes, par leur action de dégradation, vont aussi stimuler l'expression de gènes promoteurs du cancer. La digestion de la z-Catherine (protéine intercellulaire) va provoquer l'activation de la caténane (protéine intracellulaire) et sa migration vers le noyau où elle va promouvoir la division cellulaire. La β -caténane et la c-Catherine font partie des protéines d'adhésion cellulaire qui participent au comportement de la cellule et notamment au phénomène d'inhibition de contact. Ce processus empêche la cellule de proliférer lorsqu'elle est en contact avec une autre cellule. Cette perte d'adhérence cellulaire prépare alors le tissu pour l'invasion.

2.5. Classification des cancers

D'après Hontaas [38], les différents types de cancers ont été répertoriés et classés en fonction du type cellulaire à l'origine de la tumeur qui se développe (classification anatomopathologique).

Types de cancers	Dénomination en fonction du type cellulaire		Exemples
Cancers « solides »	Carcinomes	Cellules épithéliales	Cancer du sein, poumon, prostate
	Sarcome	Cellules mésenchymateuses, musculaires, osseuses	Cancer des os, du cartilage
	Cancers Neuro ectodermiques	Cellules nerveuses	Tumeurs astrocytaires, méningées
Cancers hématopoïétiques ou « liquides »	Leucémies	Précurseurs des cellules sanguines	Cancer de la moelle osseuse
	Lymphomes	Lymphocytes	Cancer du système lymphatique

Tableau 1 Présentation des différents de tumeur en fonction du tissu d'origine

2.5.1. La classification TNM des cancers

Le TNM est un système de classification des tumeurs cancéreuses solides de l'Union internationale contre le cancer (UICC). C'est le système de classification le plus fréquemment utilisé en oncologie.

T : La taille de la tumeur, la classification se fait de T0 à T4

N : Pour l'envahissement ganglionnaire, de N0 à N3, N0 correspondant à l'absence d'envahissement.

M : pour la présence de métastase, M0 à M1, M0 correspondant à l'absence de métastases, et

M1 : correspond à la présence de métastases[39].

2.6. Type de cancer

2.6.1. Cancer du sein

Chez la femme, le sein est une glande sécrétrice hypertrophiée. Durant la lactation, des amas de lobes en forme de grains de raisin (glande mammaire ou acini) produisent le lait qui remplit et enfle un système de canalisations particulièrement ramifiées. Environ 90% des cancers du sein trouvent leur origine dans les cellules épithéliales des canaux galactophores [40].

Les cancers du sein se définissent comme une prolifération maligne de cellules épithéliales bordant les canaux ou les lobules du sein appelés respectivement carcinomes canaux ou lobulaires. S'il y a une effraction de la membrane basale, séparant l'épithélium du tissu conjonctif, ces carcinomes sont dits infiltrant. Dans le cas contraire, ils sont dits carcinomes in situ ou intra-canaux. Ils possèdent ou non un potentiel métastatique[41].

2.6.1.1. Métastases mammaires

De nombreuses cellules spécialisées différentes, incluant les fibroblastes, les cellules immunitaires, les cellules endothéliales et les cellules murales des vaisseaux sanguins et lymphatiques, avec la matrice extracellulaire, doivent former un microenvironnement approprié, nécessaire à la croissance tumorale ainsi qu'à la progression métastatique. La dissémination métastatique du cancer du sein se fait principalement via les néovaisseaux lymphatiques, dans ce cas les chaînes ganglionnaires sont rapidement envahies avec des localisations au niveau des ganglions axillaires, mammaires internes, susclaviculaires et ceux de la région du cou, et un envahissement régional peut aboutir à l'atteinte de la paroi thoracique. Ceci provoque ainsi l'apparition des foyers secondaires au niveau des os, des poumons, du foie, du sein controlatéral, ou encore du cerveau [42].

2.6.2. Cancer de la prostate

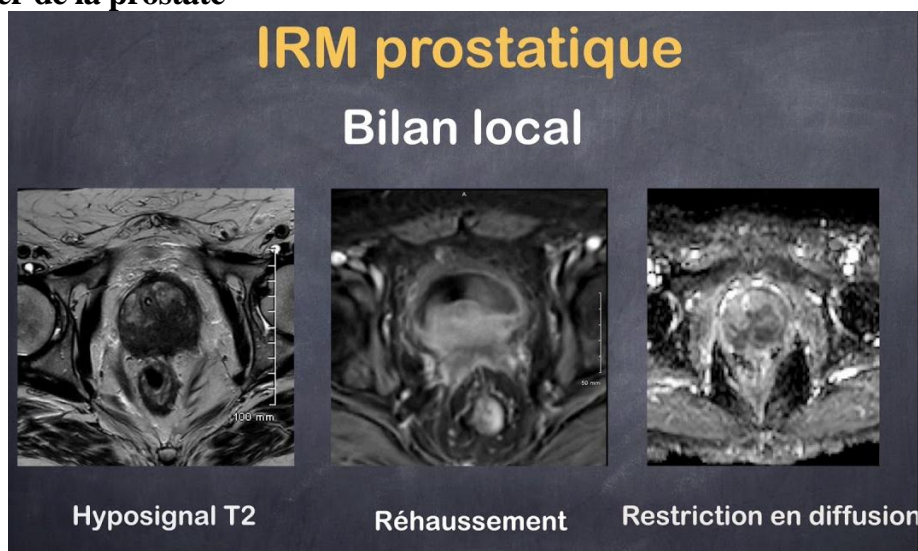


Figure 1 Figure Cancer de la prostate

Le cancer de la prostate est la plus fréquente des néoplasies chez l'homme en occident. Il se manifeste exceptionnellement avant 50 ans. Son incidence est en nette progression au cours de ces dernières décennies. Cette incidence en croissance est liée au vieillissement de la population [43].

Ce type de cancer est souvent caractérisé par une prolifération incontrôlée des cellules épithéliales de la glande prostatique, qui se multiplient de façon anarchique pour former une tumeur maligne. Cette dernière peut s'étendre localement. On parle à ce stade de cancer localisé ou intracapsulaire. Ces cellules peuvent ensuite éventuellement migrer hors de la

prostate, essentiellement vers les ganglions lymphatiques et les os, entraînant des métastases. On parle alors de cancer non localisé ou extra-capsulaire [44].

Cette maladie a souvent une évolution lente sur plusieurs années. Pour la plupart des hommes atteints, l'évolution lente de la tumeur n'entraîne pas de signes cliniques ou de symptômes au cours de leur vie [44].

2.6.3. Cancer du poumon

Les cancers broncho-pulmonaires sont des tumeurs malignes développées aux dépens des structures bronchiques et/ou plus rarement du parenchyme pulmonaire. Ils peuvent être primitifs ou secondaires [46].

Selon **Frusch** [47], Le cancer du poumon est la première cause de décès par cancer au monde. Malgré les avancées thérapeutiques, le pronostic de cette pathologie reste effroyable. Tous stades confondus, le pourcentage de survie à cinq ans est de 15%. Le tabac est, de loin, le premier facteur de risque de cette pathologie. Des progrès récents ont été réalisés dans l'identification des mécanismes cellulaires et des facteurs génétiques qui favorisent l'émergence du cancer pulmonaire.

2.6.3.1. Types des cancers du poumon

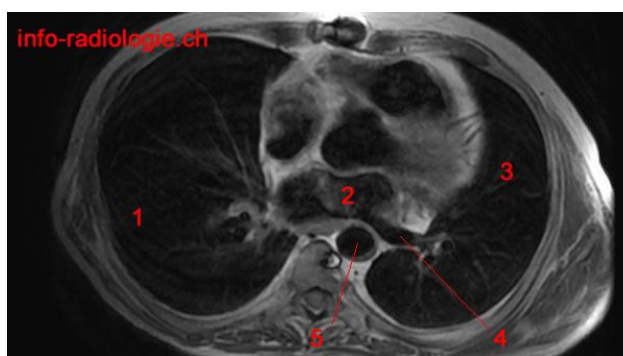
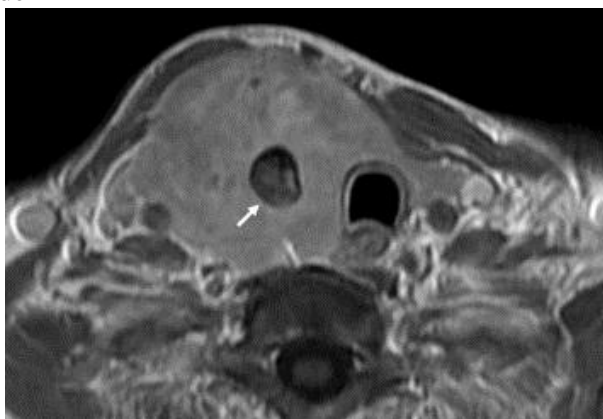


Figure 2: cancer du poumon

Plus de 90% des cancers des poumons sont dus aux quatre types histologiques suivants : l'adénocarcinome, l'épithélioma épidermoïde, le carcinome à grandes cellules et le carcinome à petites cellules [47].

2.6.4. Cancer de la thyroïde**Figure 3: Cancer de la thyroïde**

Les cancers de la thyroïde sont des tumeurs malignes qui se développent à partir des cellules épithéliales de la glande thyroïdienne [48]. Ils se présentent sous la forme d'un nodule thyroïdien qui est souvent associées à un goitre dit " multi-nodulaire ". Il pose la question du nodule toxique, qui est toujours associé à une TSH basse, et du cancer thyroïdien, qui justifie la pratique systématique d'une cytoponction à l'aiguille fine [49]. Ils sont en majorité des cancers de bon pronostic. Ils concernent principalement les femmes (76 %) et l'âge moyen au diagnostic est d'environ 50 ans [50].

Types histologique	Fréquence
Cancers différenciés de souche folliculaire/vésiculaire (80% de carcinomes papillaires 10% de carcinomes folliculaires Et 10% de cancers peu différenciés)	>90
Cancers médullaires ou à cellules C	5%
Cancers anaplasiques	Rare

Tableau 2 : Principaux types histologiques des cancer de la thyroïde[50]

D'après El-Jai, [49], les cancers thyroïdiens différenciés sont les premiers cancers malins endocriniens mais reste néanmoins rare ne représentant que 1% des pathologies néoplasiques. Ils sont représentés par :

- 1) Les carcinomes papillaires : qui sont les plus fréquents (80% de l'ensemble des carcinomes thyroïdiens) et de bon pronostic si la prise en charge est bien conduite précocement.

- 2) Les carcinomes vésiculaires ou folliculaires : qui sont beaucoup moins fréquents (5% des cancers thyroïdiens).

2.7. Diagnostique

2.7.1. Examens biologique

Bilan initial réalisé à des intervalles réguliers, numération formule sanguine (NFS), numération plaquettaire, vitesse de sédimentation (VS), la C-réactive protéine (CRP). D'autres examens biologiques apprécient le retentissement du cancer sur le fonctionnement de l'organe affecté [51].

2.7.2. Marqueurs tumoraux

Les marqueurs tumoraux sont des substances présentes dans le sang (ou éventuellement les urines) des patients cancéreux de façon anormale, et qui signent de façon plus ou moins spécifique la présence de cancer [52].

2.7.3. Examens d'imageries médicales

L'imagerie médicale consistant à produire des images du corps humain vivant et à les interpréter à des fins diagnostiques, thérapeutiques (imagerie interventionnelle) ou de surveillance de l'évolution des pathologies. L'imagerie médicale comprennent les radiographies, la tomodensitométrie appelée aussi Computer Tomography(CT) ou scanner, la fluoroscopie et l'angiographie. L'échographie et l'Imagerie par Résonance Magnétique (IRM) [53].

2.7.3.1. Endoscopie

L'endoscopie est une technique d'exploration et d'imagerie médicale réalisée à l'aide d'un instrument appelé endoscope. Ce dispositif médical (DM) est un « tube optique muni d'un dispositif d'éclairage, destiné à être introduit dans une cavité du corps humain pour l'examiner» [54].

2.7.4. Examens histologique

L'examen histologique permet d'établir le diagnostic de malignité à partir d'une biopsie ou d'une pièce opératoire. L'examen macroscopique, sur pièce opératoire, permet de mesurer la taille tumorale, d'étudier les berges d'exérèse et donne une orientation de diagnostic en fonction de l'aspect de la lésion. L'examen microscopique, sur biopsie ou pièce opératoire, permet de donner un type histologique et d'identifier plusieurs facteurs pronostiques (comme la différenciation et l'activité mitotique) et prédictifs (comme la recherche des récepteurs hormonaux) de la tumeur [45].

2.8. Le cancer en Algérie

Le cancer figure parmi les principales causes de morbidité et de mortalité dans le monde. Il représente à lui seul le plus important obstacle à l'augmentation de l'espérance de vie dans tous les pays du monde au 21ème siècle [61].

Le Centre International de Recherche sur le Cancer (CIRC), estime qu'un homme sur cinq et une femme sur six dans le monde développeront un cancer au cours de leur vie, et qu'un homme sur huit et une femme sur onze mourront de leur maladie [2].

Les nouvelles données mondiales sur le cancer suggèrent que le fardeau mondial du cancer a atteint 18,1 millions de cas et 9,6 millions de décès par cancer [2].

L'Algérie est un exemple de véritable transition épidémiologique [3-7]. L'entraîné un vieillissement progressif de la population, surtout chez les personnes de plus de 60 ans dans la pyramide des âges.

Cependant, la transformation de l'environnement, l'augmentation du tabagisme, stress, mode de vie sédentaire, urbanisation) et le changement de style de vie sont la cause de l'émergence des maladies non transmissibles (MNT).

Le cancer est devenu un véritable fardeau et un problème sociétal [8]. Le cancer en Algérie est en nette recrudescence avec des (APC) Annuel Pourcentage fluctuant entre 2 et 8 % pour les principales localisations (20). Les registres du cancer de population ont été priorisé dans le cadre du plan cancer 2015-2019 [9], et renforcés par m'arrêté N°22 du 18 février 2014 [10].

La mise en œuvre de ces outils institutionnels va nous permettre d'avoir une large couverture d'enregistrement et par conséquent des données épidémiologiques fiables, représentatives de l'ensemble de pays, très utile pour la recherche scientifique et la santé publique, avec des projections et des hypothèses de travail pour les années à venir [8].

L'analyse de l'incidence de la tendance et de la survie du cancer est un élément essentiel des processus de surveillance et d'observation de l'épidémiologie des cancers.

Notre but est de mettre en place une banque de données fiables d'incidence de tendances et de survie, avec une surveillance épidémiologique du cancer en Algérie, sur la base de registres de populations couvrant la région Est et Sud-est de l'Algérie.

2.8.1. Description des réseaux de registres de cancer

Le réseau national des registres a été créé en 2015 dans le cadre du plan cancer 2015-2019 [9], dans l'axe stratégique numéro 6 portant sur le développement du système d'information et de la communication sur les cancers. La mise en œuvre de ce réseau national est renforcée par l'arrêté N°22 du 18 février 2014 [10].

Cet arrêté institutionnalise les registres de population, avec la consolidation des registres existants et la mise en place de nouveaux registres dans l'ensemble des wilayas du pays. Cette institutionnalisation s'articule sous forme de coordination régionale Est, Centre et Ouest [11], pour permettre avoir une large couverture d'enregistrement du cancer en Algérie.

Les 48 wilayas du pays sont réparties dans les trois réseaux régionaux Est et Sud-est, Centre et Sud Centre, et Ouest et Sud-ouest, coordonnés par le réseau National des Registres de Cancer [10-11].

- ❖ Réseau Est et Sud-est coordonné par le Registre de Cancer de Sétif et comporte 20 wilayas : Sétif, Annaba, Bejaia, Taref, Batna, Khenchla, Skikda, Souk-Ahras, Biskra, Constantine, Om El Bouaghi, El Oued, Guelma, BBA, Tébessa, Jijel, Msila, Mila, Illizi, Ouargla.
- ❖ Réseau Centre et Sud-centre coordonné par INSP et comporte 13 wilayas : Alger, Blida, Médéa, Tipaza, Ghardaïa, Tizi-Ouzou, Djelfa, Tamanrasset, Ain Defla, Boumerdes, Bouira, Laghouat, Chlef.
- ❖ Réseau Ouest et Sud-ouest coordonné par le Registre de Cancer d'Oran et comporte 15 wilayas : Oran, Mostaganem, Tlemcen, Adrar, Relizane, Tiaret, Mascara, Tissemsilt, Sidi-Bel-Abbès, Tindouf, Naama, Saida, Bechar, Ain Témouchent, El Bayed.

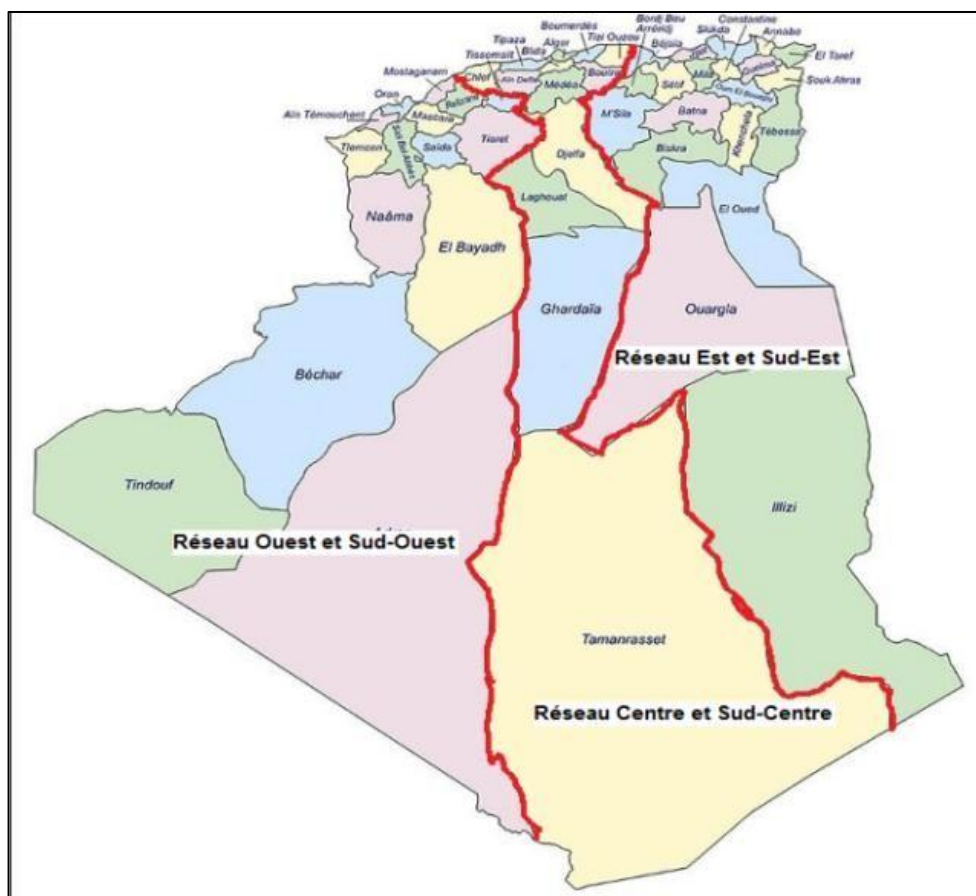


Figure 4: Les 03 Réseaux Régionaux des Registres de Cancer

Source : Algerian Journal of Health Sciences Numéro Spécial 2020(Oran)

<https://ajhs.atrss.dz>

La surveillance des cancers en Algérie a déjà une histoire. Le premier registre des cancers était basé sur les données des laboratoires de Pathologie.

Il a été mis en place en 1970 par Abdenour Yaker, pathologiste. Il a été suivi du registre des cancers digestif d'Alger, registre de population de Larbi Abid mis en place avec l'aide du docteur Coleman Centre International de Recherche sur le Cancer.

Entre la fin des années 80 et le début des années 90 ont vu le jour les premiers registres généraux de population centrés sur Alger, Sétif, Oran, Sidi-Bel-Abbes, Batna, Constantine et Tlemcen. Aujourd'hui et depuis un arrêté ministériel pris par le ministère de la santé en 2014, des registres de wilayas ont été mis en place avec des fortunes diverses.

Nouveaux cas	Hommes	Femmes	Total
2015	16.748	25.122	41.870
2016	18.747	23.304	42.051
2017	18.883	25.037	43.920
Total	54.378	73.463	127.841

Tableau 3: Nouveaux cas de cancers enregistrés en Algérie de 2015 à 2017

Nouveaux cas	Hommes	Femmes	Total
Réseau est	8175	10858	19 033
Réseau centre	7 458	9 685	17 143
Réseau ouest	3250	4494	7744
Total	18 883	25037	43 920

Tableau 4 : Nouveaux cas de cancers survenus en Algérie en 2017

2.9. Traitement

2.9.1 Chirurgie

La chirurgie spécifique rassemble « tous les actes opératoires dirigés sur une tumeur cancéreuse pour pratiquer une ablation, et si possible guérir le malade ». Son rôle est aussi bien diagnostique que pronostique et thérapeutique à l'encontre d'une tumeur primitive, de ganglions lymphatiques satellites, d'une récurrence ou même de métastases [55].

2.9.2. Radiothérapie

La radiothérapie est un traitement qui utilise des rayonnements ionisants (photons et/ou électrons), c'est-à-dire un rayonnement capable de déposer assez d'énergie dans la matière qu'il traverse pour créer une ionisation (enlever ou ajouter des charges à un atome ou à une molécule) [56].

Elle est proposée en fonction du type de cancer, de son stade d'évolution et de l'état général du patient. Selon **Guennouni** [57]. ce traitement peut être :

- **Radiothérapie curative:** détruire la totalité des cellules cancéreuses.
- **Radiothérapie palliative (symptomatique) :** freiner l'évolution d'une tumeur, afin de diminuer les douleurs ou les symptômes.

2.9.3. Traitement médicamenteuses

2.9.3.1. Chimiothérapie

Contrairement à la chirurgie ou à la radiothérapie qui sont des traitements locaux, la chimiothérapie utilise un ou plusieurs médicaments qui seront diffusés dans l'ensemble de l'organisme. Selon la situation, elle est administrée avant ou, le plus souvent, après la chirurgie [56].

D'après **Guennouni** [57], la chimiothérapie consiste à administrer au malade un médicament cytotoxique destiné à la destruction des cellules cancéreuses que les investigations n'auraient pas pu détecter.

2.9.3.2. Immunothérapie

Les cellules cancéreuses mettent en œuvre différents mécanismes pour échapper aux effecteurs de la défense. L'immunothérapie a donc pour but de stimuler les défenses immunitaires pour qu'elles puissent à nouveau agir contre les cellules tumorales et les détruire. Ce traitement présente l'avantage d'épargner les cellules saines. L'immunothérapie consiste soit à :

- ✓ **L'immunothérapie passive** : où sont administrés des effecteurs tels que des anticorps monoclonaux spécifiques ou des lymphocytes.
- ✓ **L'immunothérapie active** : qui repose sur le principe de la vaccination en administrant un antigène pour induire une réponse immunitaire [38].

2.9.3.3. Hormonothérapie

L'hormonothérapie est une méthode qui consiste à bloquer la production ou l'action de certaines hormones naturellement produites par l'organisme, qui sont connues pour favoriser la croissance du cancer. Il existe deux cancers particulièrement sensibles à l'action des hormones, le cancer du sein et le cancer de la prostate [39].

Les hormones féminines sont l'œstradiol et la progestérone tandis que l'hormone masculine est la testostérone. Dans les deux cas il existe différentes possibilités de bloquer l'effet des hormones sur leurs récepteurs tumoraux. De leur production à partir de l'axe hypothalamo-hypophysaire jusqu'à leur fixation au niveau des récepteurs, l'arsenal thérapeutique dispose de plusieurs molécules appartenant à des classes thérapeutiques différentes. L'hormonothérapie n'est pas un traitement curatif des cancers hormonodépendants, elle permet cependant de ralentir l'évolution de la maladie [38].

2.9.3.4 Thérapie ciblée

Le terme de «thérapie ciblée» est utilisé pour désigner certaines des nouvelles molécules, disponibles entre autres dans l'arsenal thérapeutique oncologique, qui interfèrent de façon relativement ciblée avec la biologie cellulaire. Les thérapeutiques ciblées, par un mécanisme non directement cytotoxique, visent à contrôler la maladie sur une longue période. Selon leur nature et leur mode d'action, ces molécules vont s'intégrer dans une stratégie thérapeutique globale où elles feront partie de schémas utilisant conjointement la chimiothérapie et/ou l'hormonothérapie et/ou la radiothérapie ; Leur mode d'administration simplifié et leur profil d'effets secondaires en fait des thérapies rapidement adoptées[58].

2.10 Plan national cancer 2015-2019

Le plan est la suite logique du résultat des deux rapports d'évaluation et de suivi du plan cancer, suite à la mission confiée au Professeur Messaoud Zitouni par Monsieur le Président de la République pour «assurer, en liaison avec les autorités concernées, le suivi et l'évaluation du plan national cancer et proposer toutes mesures utiles à l'effet d'améliorer davantage la qualité des soins et du suivi des patients».

L'idée du "Plan National Cancer 2015-2019" était déjà suggérée dans le rapport préliminaire qui concluait : "Sur la base des recommandations des groupes de réflexion il sera nécessaire d'élaborer un plan de lutte contre le cancer rédigé, organisé, consensuel et surtout structuré autour d'objectifs clairs, précis, ambitieux tout en étant réalistes et réalisables avec un échéancier raisonnable entre cinq et dix ans". La réflexion sur le sujet s'est progressivement développée lors de l'élaboration du rapport d'octobre 2013.

Le Plan National Cancer 2015-2019 s'est largement inspiré des propositions et recommandations de ces deux précédents rapports.

2.10.1 Objectif principal du plan

"La réduction de la mortalité et de la morbidité par cancer et l'amélioration de la qualité de vie des malades pendant et après le traitement", de même qu'il a fixé les huit axes stratégiques qui ont été hiérarchisés selon un ordre prioritaire avec pour chaque axe une cible principale ou "focus".

Cette hiérarchisation a été établie sur le critère du degré d'efficacité du résultat attendu pour le malade.

Les cibles principales ont été choisies, soit à cause de leur importance soit à cause du caractère d'urgence de leur réalisation. Enfin l'identification de certains dysfonctionnements,

signalés très souvent par les experts, nous a poussés à insister plus particulièrement sur des thèmes méritant une attention particulière que sont certains focus.

L'axe Stratégique 1 concerne l'amélioration de la prévention contre les facteurs de risque. La prévention reste, en effet, la seule possibilité reconnue pour réduire la mortalité et l'incidence des cancers. La cible principale ou focus dans ce plan est la lutte contre le tabagisme, principale cause actuellement identifiée de cancer. Un Plan Stratégique National Multisectoriel de Lutte Intégrée contre les Facteurs de Risque des MNT 2014-2018, élaboré par la DGPPS du MSPRH, comprend les principales mesures contre le tabagisme et pourrait représenter la feuille de route de cette lutte pour les 5 prochaines années.

L'axe Stratégique 2 concerne l'amélioration du dépistage de certains cancers. Le focus a été mis sur le dépistage du cancer du sein dont l'augmentation très forte devient particulièrement inquiétante. Une réorganisation structurée du dépistage de ce cancer, associée à une meilleure rentabilisation des ressources matérielles et humaines existantes, s'impose dans le cadre de l'urgence. En effet, des expériences internationales de dépistage organisé et réussi ont démontré que des résultats spectaculaires pouvaient être obtenus rapidement en terme non seulement de bénéfice médical pouvant atteindre 90 % de guérison mais aussi de gain économique. Le dépistage du cancer du col de l'utérus, cancer gynécologique moins fréquent nécessite, la redynamisation d'un programme existant mais insuffisant.

L'axe Stratégique 3 concerne l'amélioration du diagnostic du cancer. Le focus est centré sur l'AnatomoCyto-Pathologie, examen clé incontournable parce qu'il est le seul capable d'identifier avec certitude le cancer, son type et son stade et à suivre son évolution. Cette spécialité gagnera à être renforcée en l'organisant en réseaux, en renforçant les qualifications humaines et en augmentant ses ressources budgétaires. De plus, l'amélioration de ses performances permettra une meilleure prise en charge des malades en ciblant mieux les traitements et en permettant un suivi plus pertinent donc moins pénible pour le malade.

Les examens biologiques devront être enrichis par l'introduction de méthodes modernes spéciales au cancer, telles que la biologie moléculaire et la génétique qui sont à l'avant-garde des progrès dans ces domaines.

Les plateaux d'Imagerie Médicale doivent être modernisés à cause d'une double mutation qui affecte cette spécialité dans le sens du diagnostic et du traitement et dont elle devient partie prenante (radio chirurgie).

L'axe stratégique 4 concerne la redynamisation du traitement. Le focus a été mis sur l'interdisciplinarité parce qu'elle représente, de l'avis unanime des experts, le seul mode efficace et efficient dans la prise de décision thérapeutique pour le cancer. Il est urgent qu'elle soit institutionnalisée dans notre pays, non seulement sur le plan réglementaire mais aussi et surtout sur le plan comportemental de la part des professionnels.

La radiothérapie mérite une attention particulière à cause de ses spécificités : c'est le traitement le plus techno-dépendant exigeant des appareillages particulièrement sophistiqués dont la maintenance est d'une importance capitale et nécessite des personnels variés mais complémentaires. Avec l'amélioration progressive enregistrée, depuis une date récente, la radiothérapie devra rapidement rejoindre la chimiothérapie et la chirurgie dans un cadre de normalité pour les patients.

Le domaine de la chimiothérapie est lui aussi en constante mutation avec l'utilisation de nouvelles drogues, réputées de plus en plus efficaces mais coûtant aussi de plus en plus cher. C'est à ce niveau que le rôle de l'interdisciplinarité est capital, car ce n'est que dans un esprit de collégialité respectant les normes scientifiques et les règles éthiques que le traitement de tous les malades se fera avec le plus d'efficacité et le plus d'efficience.

Les traitements palliatifs permettent de faire la transition avec l'axe suivant. **L'axe stratégique 5** concerne l'organisation de ces traitements palliatifs, l'orientation, l'accompagnement et le suivi du patient. Le focus a été placé sur les "cellules d'écoute et d'orientations" dédiées à ces missions dans le cadre intégré d'une équipe multidisciplinaire. Ces cellules ont pour rôle d'assurer le soutien psychologique des patients et de leurs familles et de réduire les obstacles à l'accès aux centres de diagnostic et de traitement du cancer.

Réparties harmonieusement à travers le territoire national, elles devraient être mieux organisées notamment par la désignation d'un médecin généraliste coordonnateur en cancérologie. Ce corps de métier pourrait être rapidement créé. Dans ce cadre, le rôle des associations d'aide aux malades cancéreux est important, en maintenant le lien entre les malades, leur entourage et le milieu soignant ainsi que d'autres segments de la société

L'axe stratégique 6 vise à développer le système d'information sanitaire et de communication sur le cancer. Le focus a été mis sur les registres du cancer qui sont les instruments centraux de la surveillance épidémiologique dans ce domaine. L'Algérie dispose de Registres performants reconnus par les instances internationales. Il s'agit actuellement d'améliorer la coordination et l'extension de ces registres à travers le territoire et de renforcer les capacités d'intervention.

L'amélioration de l'information et de la communication sur le cancer n'est pas encore assez efficace, le cancer souffrant encore de tabous et de préjugés. Les efforts de communication devront concerner non seulement les malades et les citoyens mais aussi les professionnels de la santé.

L'axe stratégique 7 concerne le renforcement de la formation et de la recherche sur le cancer. La notion de nouveaux métiers, qui sont cités dans le plan, propre au domaine de la cancérologie a été prise comme focus. Le constat fait par les experts est que des métiers prenant en charge le cancer sont de plus en plus en décalage par rapport au système de formation aussi bien sur le plan quantitatif que qualitatif. Il s'agit d'optimiser la formation de tous les acteurs de la chaîne de soins dans le domaine de la cancérologie et de développer de nouvelles compétences imposées par les progrès constants dans la prise en charge des malades.

Le rôle de l'intersectorialité est capital à ce niveau. L'implication des secteurs de l'enseignement supérieur et de la formation professionnelle pour la réforme de la formation et celle des finances et de la fonction publique, pour la définition des statuts futurs des nouveaux métiers, de leur reconnaissance et de leur valorisation professionnelle est indispensable.

Le secteur de la cancérologie étant en pleine expansion et constante mutation mérite un développement plus significatif de la recherche aussi bien dans un cadre opérationnel que translationnel.

La recherche doit être revalorisée tant au niveau du Ministère de la santé que de celui de l'Université.

L'axe stratégique 8 vise à renforcer les capacités de financement de la prise en charge des cancers avec comme focus d'optimiser et de rationaliser les ressources financières disponibles. L'utilisation de celles-ci qui sont importantes, devra être optimisée et rationalisée. En effet, l'ouverture de nombreux hôpitaux, CHU et surtout de Centres Anti Cancer avec des équipements très modernes vont entraîner une mobilisation substantielle des ressources financières.

Les domaines de l'information, de la prévention, du dépistage, de l'organisation des soins, de la formation et de la recherche ont été jusqu'à présent défavorisés par rapport au secteur des soins et des infrastructures. Il est nécessaire de revoir la budgétisation de leurs programmes parce qu'ils portent eux, les véritables sources du succès des actions de ce plan.

3. Propositions de mesures

Les mesures d'urgence proposées concernent essentiellement six domaines :

3.1- L'amélioration de l'accueil du patient

et le diagnostic précoce en réorganisant la chaîne des soins à partir du lieu de domicile du malade. Dans ce domaine le rôle des médecins généralistes exerçant dans les structures publiques et /ou privées est capital. Ce rôle peut, en effet, être exercé dans les cellules d'accueil et d'orientation qui doivent être développées et redynamisées ou par un « médecin référent » exerçant dans le secteur privé.

3.2- La disponibilité des moyens thérapeutiques

concerne essentiellement la radiothérapie qui représente actuellement le problème le plus crucial. Les mesures déjà proposées dans le rapport préliminaire sont reprises dans ce rapport, elles méritent une attention particulière.

De plus, il est nécessaire de développer la pharmacie oncologique et d'installer des unités de reconstitution des traitements anticancéreux de façon planifiée.

3.3- La rapidité du diagnostic

peut être améliorée par l'extension sur le territoire national et l'équipement des laboratoires d'Anatomo-cyto-pathologie ainsi que par l'organisation du travail en réseau.

3.4- La mise en application effective des textes réglementaires

concernant les programmes de prévention et notamment la lutte anti-tabac.

3.5- Le rôle capital et l'intérêt de l'intersectorialité

surtout dans les domaines de la prévention qui ne peut être mise en place sans la synergie avec les secteurs autres que celui de la santé. En particulier la collaboration avec le Ministère de l'Enseignement et de la Recherche Scientifique ainsi que celle avec le Ministère du Travail et de la Sécurité Sociale doivent être renforcées.

3.6- L'adaptation de l'enseignement aux nouveaux métiers

concernant la prise en charge du cancer est impérative aussi bien dans le champ médical, paramédical qu'administratif. Il est d'autant plus urgent de répondre à cet impératif que cette adaptation passe par l'établissement de nouveaux programmes adaptés aux nouvelles connaissances. De plus, les grandes mutations et innovations scientifiques et technologiques et économiques sont actuellement tellement rapides et profondes qu'elles nécessitent la mise en place de programmes de recherche pour qu'à l'avenir, et en cas de besoin, l'Algérie puisse s'adapter plus rapidement.

Enfin, la mise en place d'un système d'évaluation performant représente une urgence stratégique, en ce sens que toutes les propositions mesures et actions que nous venons d'évoquer et

qui figurent dans ce rapport devront, lorsqu'elles seront mises en œuvre, être dotées d'indicateurs qui serviront d'instruments pour leur évaluation.

En conclusion, il ne fait pas de doute que cette entreprise d'amélioration de la prise en charge des patients cancéreux soulève un vif intérêt, aussi bien auprès des usagers que des personnels de santé et des associations, et induit une forte attente qui peut être résumée par une phrase saisie dans une contribution envoyée par un participant : "La réussite de ce plan est pour nous un défi, et une rude bataille que nous devons gagner, car elle sera la locomotive qui mènera tout notre système de santé vers des changements auxquels nous aspirons tous".

Conclusion

Dans ce chapitre nous avons présenté et défini les différents concepts liés au phénomène du cancer. Le cancer est une maladie fréquente dont les causes sont encore mal connues, mais qui, dans certains cas, pourrait être évitée. Cette conviction vient en particulier de l'observation de l'évolution spatio-temporelle de la maladie et de l'analyse approfondie de l'association entre certaines expositions et sa survenue. Il incombe à l'épidémiologie de rassembler les données scientifiques qui permettent de mieux comprendre le rôle des différents facteurs de risque, de quantifier leurs effets et leur synergie éventuelle, et d'utiliser cette connaissance pour organiser la prévention, étude de la fréquence des maladies dans les populations (nombre de cas, nombre de décès).

L'Algérie, comme le reste des pays du monde, se trouve confrontée à ce fardeau, elle présente plusieurs facteurs de risque augmentant l'incidence et la mortalité du cancer à un rythme inquiétant. A travers les années, elle alloué des budgets importants dans la tentative de diminuer l'impact de cette maladie sur la population.

CHAPITRE III

Conception

d'application

Introduction

Dans le cadre de la conception d'application nous présenter fonction de l'application

1. Spécification informelle du banc de test anapath

La spécification informelle est un type de cahier de charge pour exprimer les besoin de fonctionnement de ce système on présentant une description de l'architecture matérielle du système et une description des fonctions du système et les paramètres utilisé.

1.1 Aspect matériel

Base de données d'instruments :

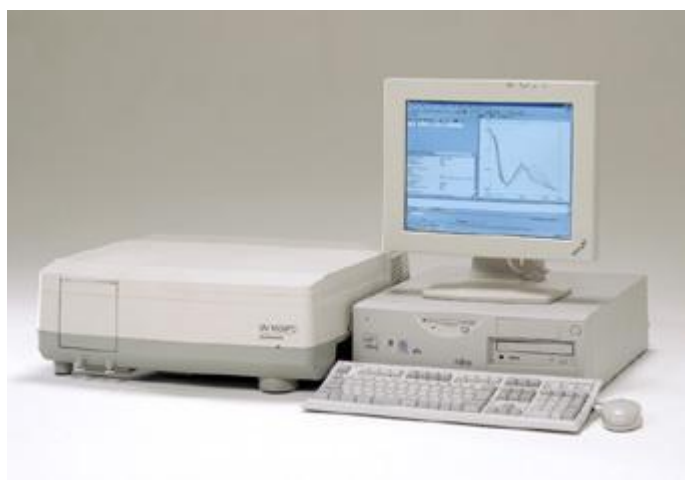


Figure 5 : Shimadzu Europe-UV-1650PC

L'UV-1650PC combine des performances optiques avec une UVProbe conforme aux BPL/BPF - le logiciel le plus avancé disponible.

Le banc optique UV-1650PC ne ressemble à aucun autre système sur le marché. Ses optiques scellées hautes performances sont logées dans un boîtier résistant aux chocs. Cela signifie une durée de vie prolongée pour votre instrument et une grande durabilité pour la plupart des solvants. La conception à double faisceau offre le meilleur de la technologie de réseau moderne avec le monochromateur monté sur un banc optique très stable pour assurer l'intégrité et un fonctionnement sans problème pour les années à venir. Une amélioration supplémentaire de la stabilité à long terme est obtenue grâce à la surveillance et à l'ajustement automatiques des fluctuations de la puissance de la lampe et de l'électronique du système. Le réseau holographique blazer avec des lampes au deutérium et tungstène-halogène auto-alignantes et optimisant l'énergie améliore la précision sur tout le spectre UV-VIS et dans le NIR de 190 à 1100 nm. Les données

spectrales sont obtenues à bande passante constante avec une résolution inférieure à 2 nm. La superbe attention portée aux détails dans la conception de cet instrument vous offre toutes ces fonctionnalités tout en conservant un faisceau ultra-étroit de seulement 1 mm, vous permettant d'examiner vos plus petits échantillons en toute confiance. Les caractéristiques de conception exclusives s'adaptent à n'importe quelle taille ou type d'échantillon. Vous pouvez rapidement passer d'une cuvette standard à une cellule capillaire ou à un support de poudre ou de film. Avec une variété complète de fonctionnalité standard et un large choix d'options, l'UV-1650PC répond aux exigences de la plupart des industries et des applications. Vous pouvez rapidement passer d'une cuvette standard à une cellule capillaire ou à un support de poudre ou de film. Avec une variété complète de fonctionnalité standard et un large choix d'options, l'UV-1650PC répond aux exigences de la plupart des industries et des applications. Vous pouvez rapidement passer d'une cuvette standard à une cellule capillaire ou à un support de poudre ou de film. Avec une variété complète de fonctionnalité standard et un large choix d'options, l'UV-1650PC répond aux exigences de la plupart des industries et des applications.

Aux États-Unis et en Europe, UV-1650PC est disponible en tant que banc optique UV1601PC avec UVProbe.

Shimadzu a maintenant fait passer la spectroscopie au niveau supérieur avec le logiciel UVProbe puissant, flexible et facile à utiliser. Ce package est spécialement conçu pour Windows NT. Ses modes d'acquisition de données standard, ses vastes capacités de traitement de données et son générateur de rapports robuste constituent l'outil ultime pour vos solutions UV-Vis.

Ce package 32 bits, à utiliser avec les spectrophotomètres UV-Vis des séries UV-1600, 2400, 2500 et 3100, a été conçu pour s'adapter à tous les niveaux de fonctionnement, des utilisateurs débutants aux utilisateurs avancés. Le logiciel a la capacité d'afficher ou de masquer n'importe quel élément de l'écran pour répondre aux besoins de fonctionnement essentiels des utilisateurs pour leur application spécifique. La fonction de sécurité permet à l'administrateur de déterminer les opérations qu'un groupe d'utilisateurs particulier peut effectuer et enregistre le nom de l'utilisateur ainsi que l'opération dans l'historique de l'ensemble de données.

Les capacités de traitement et le générateur de rapports robuste fournissent l'outil ultime pour vos solutions UV-Vis.

Caractéristiques

Plage de longueurs d'onde : 190,0 à 1100,0 nm

Bande passante spectrale : 2 nm

Affichage de la longueur d'onde/ Lisible jusqu'à 0,1 nm Réglage de la longueur d'onde

Incréments de 0,1nm (incréments de 1 nm pour le balayage de la longueur d'onde)

Précision de la longueur d'onde : $\pm 0,5$ nm (correction automatique de la longueur d'onde)

Reproductibilité de la longueur d'onde : $\pm 0,1$ nm Vitesse de balayage de la longueur d'onde
Environ 6 000 nm/min.

Vitesse de balayage de la longueur d'onde Environ 3200nm/min. à environ 160 nm/min.

Intervalle de regroupement des données Sélection automatique de 2,0, 1,0, 0,5, 0,2, 0,1 nm

Longueur d'onde d'échange de lampe 295,0 ~ 364,0 nm (340,8 nm)

Lumière parasite : Moins de 0,05 % (à 220,0 nm et 340,0 nm)

Système photométrique : Optique à double faisceau

Plage photométrique : Absorbance : -0,5 ~ 3,999 Abs.

Transmittance : 0,0 ~ 300 %

Précision photométrique $\pm 0,004$ abs. À 1,0 Abs. (Testé avec le filtre NIST 930D) $\pm 0,002$
Abs. À 0,5 Abs.

Reproductibilité photométrique $\pm 0,002$ Abs. À 1,0 Abs / $\pm 0,002$ Abs. À 0,5 Abs.

Stabilité de la ligne de base : Moins de $\pm 0,001$ Abs. /heure

Planéité de la ligne de base : $\pm 0,002$ Abs.

Correction de la ligne de base : Automatique avec mémoire d'ordinateur, en deux étapes
grossière et fine

Source lumineuse : Lampe halogène de 50 W (durée de vie de 2 000 heures) et lampe au
deutérium

Monochromateur : Réseau holographique blazer concave corrigé des aberrations

Détecteur : Photodiode au silicium

Compartiment échantillon : Dimensions intérieures : L110,0 X P230,0 X H105,0 mm (en
partie 105 mm de profondeur)

Distance entre les faisceaux lumineux : 100,0 nm

Installation : Fixé avec 2 vis

Taille du faisceau : 10 X 1 mm

Interface ports : Port RS-232C f

Alimentation requise : 100, 120, 220, 240 VCA, 50/60 Hz, 160 VA.

Conditions ambiantes Température : 15~35°C

Humidité : 45~80%

Dimensions et poids : W550 X D470 X H200mm, 18kg

2. Spectrophotomètre UV-VIS V-630 [59]

Analyse de routine grâce à des méthodes de recherche avancées



Figure 6 : Spectrophotomètre UV-VIS V-630

Le V-630 est un instrument économique à usage général avec une conception compacte pour économiser de l'espace sur la paillasse. Ses excellentes performances spectroscopiques conviennent aux applications de routine, de CQ et de recherche. Les fonctionnalités conviviales incluent la fonction d'accessoire IQ pour la reconnaissance automatique des accessoires et le démarrage IQ pour le démarrage immédiat des applications de collecte de données enregistrées lors de la réalisation de mesures de routine. Deux interfaces utilisateur graphiques sont disponibles, notamment un module distant intelligent (IRM) nouvellement repensé avec un écran tactile LCD couleur ou le logiciel Spectra Manager™ II, la dernière version du logiciel innovant de spectroscopie multiplateforme de JASCO. Ces deux interfaces de contrôle et d'analyse permettent un contrôle complet du système et un traitement avancé des données. Spectra Manager CFR est une version de logiciel conforme à la norme 21 CFR partie 11 et est disponible en option.

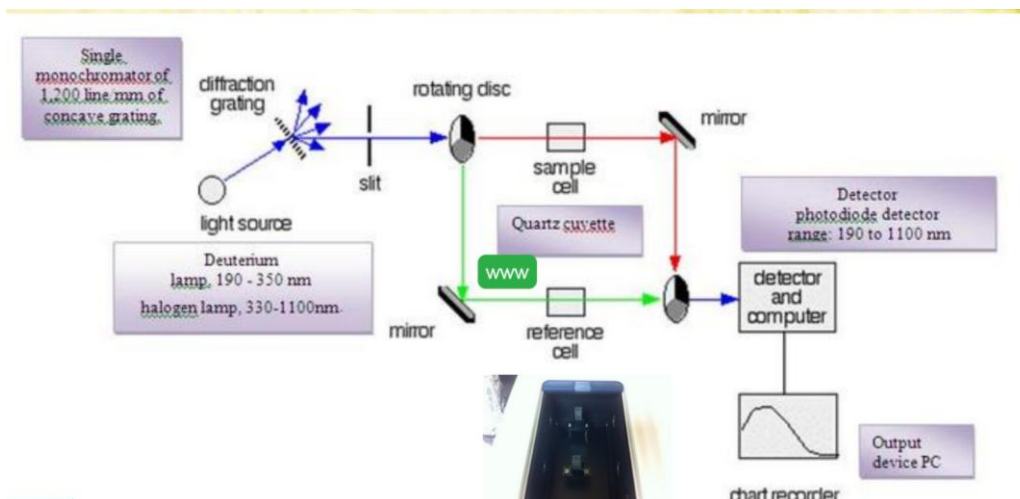


Figure 7 : Fonction Spectrophotomètre UV-VIS V-630

- **Simplicité et facilité d'utilisation**

Les logiciels IRM et Spectra Manager II offrent une interface intuitive permettant même aux utilisateurs novices de collecter des données en toute confiance.

- **Conception compacte**

Le banc optique du V-630 est un compact de 441 (L) x 468 (P) x 216 (H) mm permettant de gagner de la place dans un laboratoire moderne.

- **Balayage à grande vitesse**

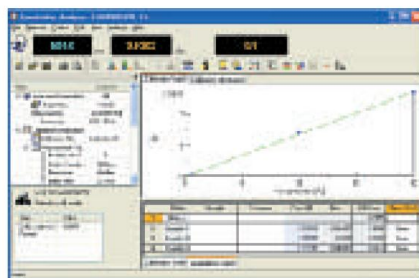
Des optiques à haut débit et des détecteurs à réponse rapide permettent au V-630 de balayer à des vitesses allant jusqu'à 8 000 nm/min. sans erreurs de suivi de longueur d'onde.

- **Large gamme d'accessoires d'échantillonnage**

Une gamme complète d'accessoires d'échantillonnage, notamment des porte-cuves, des cuves à circulation, des contrôleurs de température et des progiciels, est disponible pour optimiser le V-630 pour des applications spécifiques.

- **Validation des instruments**

Pour les laboratoires conformes aux réglementations GMP/GLP, le V-630 offre une routine de validation des instruments (en standard) pour vérifier les performances des instruments.



Spectra Manager II



iRM-700

Figure 8 : Affichage Spectrophotomètre UV-VIS V-630 [60]

3. Diffractomètre à rayons X (Rigaku MiniFlex 600)

**Figure 9 : Rigaku MiniFlex 600**

Aperçu

Le Miniflex 600 est un diffractomètre à rayons X à usage général de paillasse qui peut effectuer une analyse qualitative et quantitative des matériaux polycristallins. C'est un outil essentiel dans l'identification et la quantification des phases et la caractérisation de la structure cristalline pour un large éventail d'applications telles que les matériaux, la géologie, la médecine légale, l'alimentation, la pharmacie et les études environnementales.

Principe d'opération

Le rayon X est généré dans un tube à rayons X Cu, filtré par un monochromateur pour augmenter la sensibilité, collimaté pour se concentrer et est dirigé vers l'échantillon. Le rayon X est diffracté par les couches atomiques du cristal, et le diagramme de diffraction est utilisé pour

l'analyse basée sur la loi de diffraction de Bragg ($n\lambda=2d \sin\theta$). Les positions 2θ des pics de diffraction fournissent une empreinte digitale unique des phases présentes et l'intensité et la propagation des pics sont utilisées pour obtenir la quantité de chaque phase sur la base de la méthode d'analyse quantitative de Rietveld.

Composants

Le système est équipé d'un tube à rayons X Cu de 2,0 kW, d'un détecteur de compteur à scintillation NaI, d'un monochromateur en graphite et d'un passeur d'échantillons automatisé à 6 positions avec centrifugeuse d'échantillons. Le contrôle de l'instrument et le traitement des données sont effectués à l'aide d'un PC. Le progiciel d'analyse complet PDXL de Rigaku est utilisé pour la réduction des données, l'élimination du bruit de fond, la recherche et la correspondance des pics avec les bases de données ICDD PDF, la détermination de la taille des cristallites, la détermination du % de cristallinité et l'analyse quantitative.

Fabricant Rigaku

Modèle MiniFlex 600

Catégorie Caractérisation des matériaux

Application

Le diffractomètre à rayons X de paillasse MiniFlex est un instrument de diffraction de poudre polyvalent qui peut être utilisé pour :

- identification des phases
- quantification de phase (identification de phase)
- pourcentage (%) cristallinité
- taille et déformation des cristallites
- raffinement des paramètres de réseau
- raffinement rietveld
- structure moléculaire

La description

Détecteur à matrice de pixels hybride HyPix-400 MF 2D (HPAD) [61]

Passeur d'échantillons automatique à 8 positions

Logiciel:

- contrôle des instruments : contrôle et mesure
- analyse des données : PDXL

Générateur de rayons X :

- puissance maximale : 600 W
- tension des tubes : 40 kV
- courant du tube : 15 mA

Volet : volet roulant lié au verrouillage

Tube à rayons X : Cu

Optique:

- fente de divergence : fixe ou variable
- fente de diffusion : fixe
- fente de réception : fixe
- filtre : filtre à feuille $K\beta$
- monochromateur : graphite
- fente de Soller : $5,0^\circ$ ou $2,5^\circ$

Goniomètre:

- type : verticale
- rayon : 150 mm
- plage de balayage : -3 à 145° (2θ)
- vitesse de balayage : $0,01$ à $100^\circ/\text{min}$ (2θ)
- largeur de pas minimale : $0,005^\circ$ (2θ)
- précision : $\pm 0,02^\circ$

Détecteur:

- compteur à scintillation : scintillateur NaI
- D/teX Ultra : détecteur de bande de silicium à grande vitesse

4. Structure de réseau

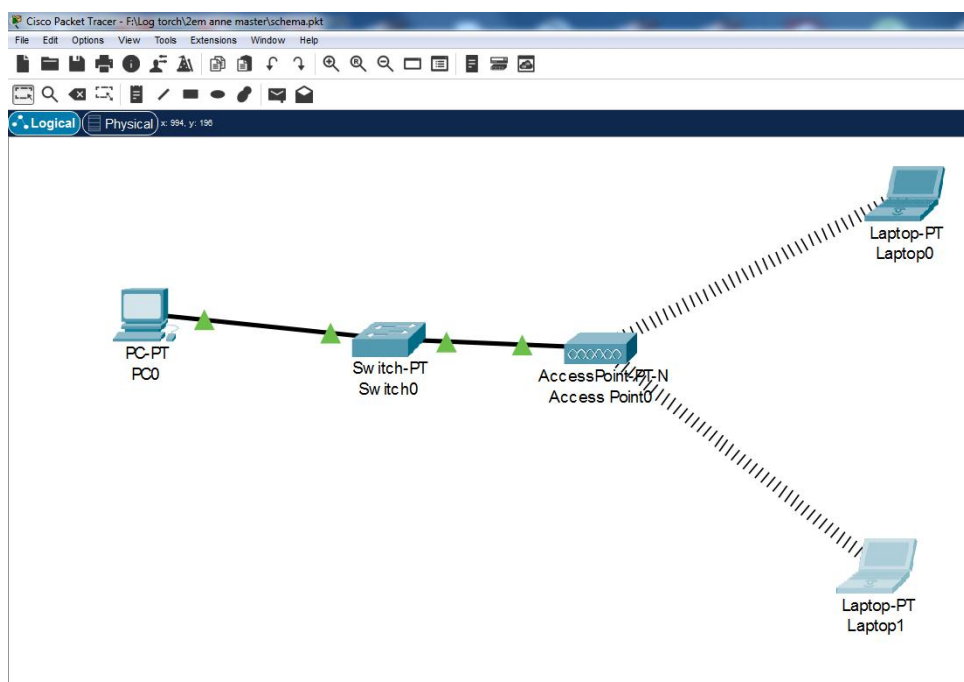


Figure 10: Installation de réseau (Schéma tracer par Cisco Packet Tracer)

5. Conception

Il y a plusieurs des méthodes de conception (UML, Merise ...), dans notre projet on a utilisé méthodes UML (Unified Modelling Language) :

Le Langage de Modélisation Unifié, est un langage de modélisation graphique à base de pictogrammes conçu comme une méthode normalisée de visualisation dans les domaines du développement logiciel et en conception orientée objet.

5-1. UML

Le Langage de Modélisation Unifié, de l'anglais **Unified Modeling Language** (UML), est un langage de modélisation graphique à base de pictogrammes conçu comme une méthode normalisée de visualisation dans les domaines du développement logiciel et en conception orientée objet.[62]

L'UML est une synthèse de langages de modélisation objet antérieurs : Booch, OMT, OOSE. Principalement issu des travaux de Grady Booch, James Rumbaugh et Ivar Jacobson, UML est à présent un standard adopté par l'Object Management Group (OMG). UML 1.0 a été normalisé en janvier 1997; UML 2.0 a été adopté par l'OMG en juillet 2005. La dernière version de la spécification validée par l'OMG est UML 2.5.1 (2017).

5.1.1 Utilisation

UML est destiné à faciliter la conception des documents nécessaires au développement d'un logiciel orienté objet, comme standard de modélisation de l'architecture logicielle. Les différents éléments représentables sont :

- Activité d'un objet/logiciel
- Acteurs
- Processus
- Schéma de base de données
- Composants logiciels
- Réutilisation de composants.

Il est également possible de générer automatiquement tout ou partie du code, par exemple en langage Java, à partir des documents réalisés.

5.1.2 Formalisme

UML est un langage de modélisation. La version actuelle, UML 2.5, propose 14 types de diagrammes dont sept structurels et sept comportementaux. À titre de comparaison, UML 1.3 comportait 25 types de diagrammes.

UML n'étant pas une méthode, l'utilisation des diagrammes est laissée à l'appréciation de chacun. Le diagramme de classes est généralement considéré comme l'élément central d'UML. Des méthodes, telles que le processus unifié proposé par les créateurs originels de UML, utilisent plus systématiquement l'ensemble des diagrammes et axent l'analyse sur les cas d'utilisation (« use case ») pour développer par itérations successives un modèle d'analyse, un modèle de conception, et d'autres modèles. D'autres approches se contentent de modéliser seulement partiellement un système, par exemple certaines parties critiques qui sont difficiles à déduire du code.

- UML se décompose en plusieurs parties :
 - Les vues : ce sont les observables du système. Elles décrivent le système d'un point de vue donné, qui peut être organisationnel, dynamique, temporel, architectural, géographique, logique, etc. En combinant toutes ces vues, il est possible de définir (ou retrouver) le système complet.
- Les diagrammes : ce sont des ensembles d'éléments graphiques. Ils décrivent le contenu des vues, qui sont des notions abstraites. Ils peuvent faire partie de plusieurs vues.
- Les modèles d'élément : ce sont les éléments graphiques des diagrammes.

5.1.3 Vues d'UML

Une façon de mettre en œuvre UML est de considérer différentes vues qui peuvent se superposer pour collaborer à la définition du système :

- Vue des cas d'utilisation (use-case view) : c'est la description du modèle vu par les acteurs du système. Elle correspond aux besoins attendus par chaque acteur (c'est le quoi et le qui).
- Vue logique (logical view) : c'est la définition du système vu de l'intérieur. Elle explique comment peuvent être satisfaits les besoins des acteurs (c'est le comment).
- Vue d'implémentation (implementation view) : cette vue définit les dépendances entre les modules.
- Vue des processus (process view) : c'est la vue temporelle et technique, qui met en œuvre les notions de tâches concurrentes, stimuli, contrôle, synchronisation...
- Vue de déploiement (deployment view) : cette vue décrit la position géographique et l'architecture physique de chaque élément du système (c'est le où).

Le pourquoi n'est pas défini dans UML. [62]

En UML 2.5, les diagrammes sont représentés sous deux types de vue : d'un point de vue statique ou structurelle du domaine avec les diagramme de structure (Structure Diagrams).

D'un point de vue dynamique avec les diagrammes de comportement (Behavior Diagrams) et les diagrammes d'interactions (Interaction Diagrams).

5.1.4 Diagrammes UML.

La hiérarchie des diagrammes UML 2.0 sous forme d'un diagramme de classes.

Les diagrammes sont dépendants hiérarchiquement et se complètent, de façon à permettre la modélisation d'un projet tout au long de son cycle de vie. Il en existe quatorze depuis UML 2.3.

5.1.4.1 Diagrammes de structure ou diagrammes statiques

Les diagrammes de structure (structure diagrams) ou diagrammes statiques (static diagrams) rassemblent :

- Diagramme de classes (class diagram) : représentation des classes intervenant dans le système.
- Diagramme d'objets (object diagram) : représentation des instances de classes (objets) utilisées dans le système.

- Diagramme de composants (component diagram) : représentation des composants du système d'un point de vue physique, tels qu'ils sont mis en œuvre (fichiers, bibliothèques, bases de données...)
- Diagramme de déploiement (deployment diagram) : représentation des éléments matériels (ordinateurs, périphériques, réseaux, systèmes de stockage...) et la manière dont les composants du système sont répartis sur ces éléments matériels et interagissent entre eux.
- Diagramme des paquets (package diagram) : représentation des dépendances entre les paquets (un paquet étant un conteneur logique permettant de regrouper et d'organiser les éléments dans le modèle UML), c'est-à-dire entre les ensembles de définitions.
- Diagramme de structure composite (composite structure diagram) : représentation sous forme de boîte blanche des relations entre composants d'une classe (depuis UML 2.x).
- Diagramme de profils (profile diagram) : spécialisation et personnalisation pour un domaine particulier d'un meta-modèle de référence d'UML (depuis UML 2.2).

5.1.4.2 Diagrammes de comportement

Les diagrammes de comportement (behavior diagrams) rassemblent :

- Diagramme des cas d'utilisation (use-case diagram) : représentation des possibilités d'interaction entre le système et les acteurs (intervenants extérieurs au système), c'est-à-dire de toutes les fonctionnalités que doit fournir le système.
- Diagramme états-transitions (state machine diagram) : représentation sous forme de machine à états finis du comportement du système ou de ses composants.
- Diagramme d'activité (activity diagram) : représentation sous forme de flux ou d'enchaînement d'activités du comportement du système ou de ses composants.

5.1.4.3 Diagrammes d'interaction ou diagrammes dynamiques

Les diagrammes d'interaction (interaction diagrams) ou diagrammes dynamiques (dynamic diagrams) rassemblent :

- Diagramme de séquence (sequence diagram) : représentation de façon séquentielle du déroulement des traitements et des interactions entre les éléments du système et/ou de ses acteurs.
- Diagramme de communication (communication diagram) : représentation de façon simplifiée d'un diagramme de séquence se concentrant sur les échanges de messages entre les objets (depuis UML 2.x).

- Diagramme global d'interaction (interaction overview diagram) : représentation des enchaînements possibles entre les scénarios préalablement identifiés sous forme de diagrammes de séquences (variante du diagramme d'activité) (depuis UML 2.x).
- Diagramme de temps (timing diagram) : représentation des variations d'une donnée au cours du temps (depuis UML 2.3).

5.1.5 Modèles d'éléments

- Un stéréotype est une marque de généralisation notée par des guillemets, montrant que l'objet est une variété d'un modèle.
- Un classeur est une annotation qui permet de regrouper des unités ayant le même comportement ou structure. Un classeur se représente par un rectangle conteneur, en traits pleins.
- Un paquet regroupe des diagrammes ou des unités.
- Chaque classe ou objet se définit précisément avec le signe « :: ». Ainsi l'identification d'une classe X en dehors de son paquet ou de son classeur sera définie par « Paquet A::Classeur B::Classe X ».

5.1.5 Modèles d'éléments de type commun

Modèles d'éléments UML.

Symbolique des modèles d'éléments :

Classe (class).

- Objet (object).
- Cas d'utilisation (use case).
- Paquet (package).
- Nœud (node).
- Fourche (fork).
- Acteur (actor).
- État (state).
- Activité (activity).
- État initial (initial state).
- État final (final state).
 - Interface (interface).
 - O←--- sens du flux de l'interface.
 - O)----- est un raccourci pour la superposition de ----→O et O←---

5.1.6 Modèles d'éléments de type relation

- Dépendance (dependency).
- Généralisation (generalisation).
- Association (association).
- Agrégation (aggregation).
- Composition (composition).
- Réalisation.
- Utilisation.

5.1.7 Autres modèles d'éléments

- Les stéréotypes peuvent dépendre du langage utilisé.
- Les archétypes.
- Les profils.

5.2.8 Normalisation et certification

UML n'est pas une norme en droit mais un simple standard « industriel » (ou norme de fait), parce que promu par l'OMG (novembre 1997) au même titre que CORBA et en raison de son succès. Depuis juillet 2005, la première version 2.x de UML est validée par l'OMG.

Par ailleurs, depuis 2003, l'OMG a mis en place un programme de certification à la pratique et la connaissance d'UML OCUP¹¹ qui recouvre trois niveaux de maîtrise.

Exemple de séquence de création des diagrammes

Diagramme	Étape du cycle en V
1. Diagramme de cas d'utilisation	Spécification, cahier des charges
2. Diagramme de séquence	
3. Diagramme d'activité (processus métiers)	
4. Diagramme d'activité (cinématique et/ou processus applicatifs)	
5. Diagramme de classes	Conception architecturale
6. Diagramme d'objets	
7. Diagramme de communication	
8. Diagramme de déploiement	
9. Diagramme de composants	

Tableau 5 : Exemple de séquence de création des diagrammes

5.2.9 Logiciels de modélisation UML

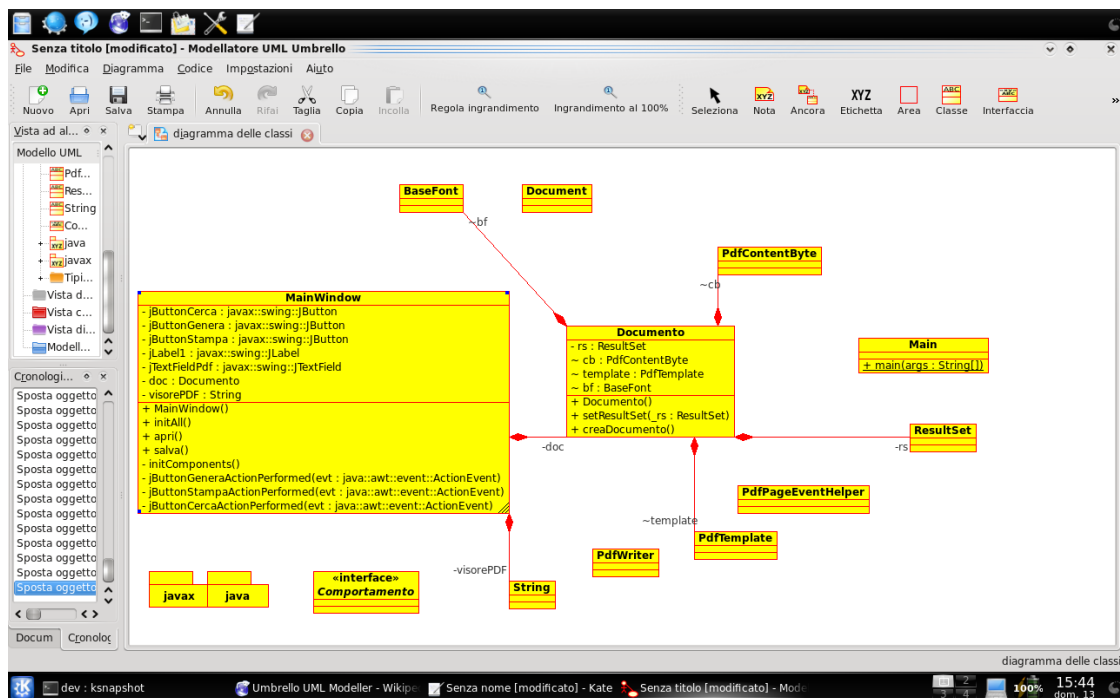


Figure 11: Modélisation d'UML

S'il existe de nombreux logiciels de modélisation UML, aucun ne respecte entièrement chacune des versions de UML, particulièrement UML 2, et beaucoup introduisent des notations non conformes. En revanche, de nombreux logiciels comportent des modules de génération de code, particulièrement à partir du diagramme de classes, qui est celui qui se prête le mieux à une telle automatisation.

Dans notre on a pris trois diagrammes :

5.3. Diagramme de cas d'utilisation

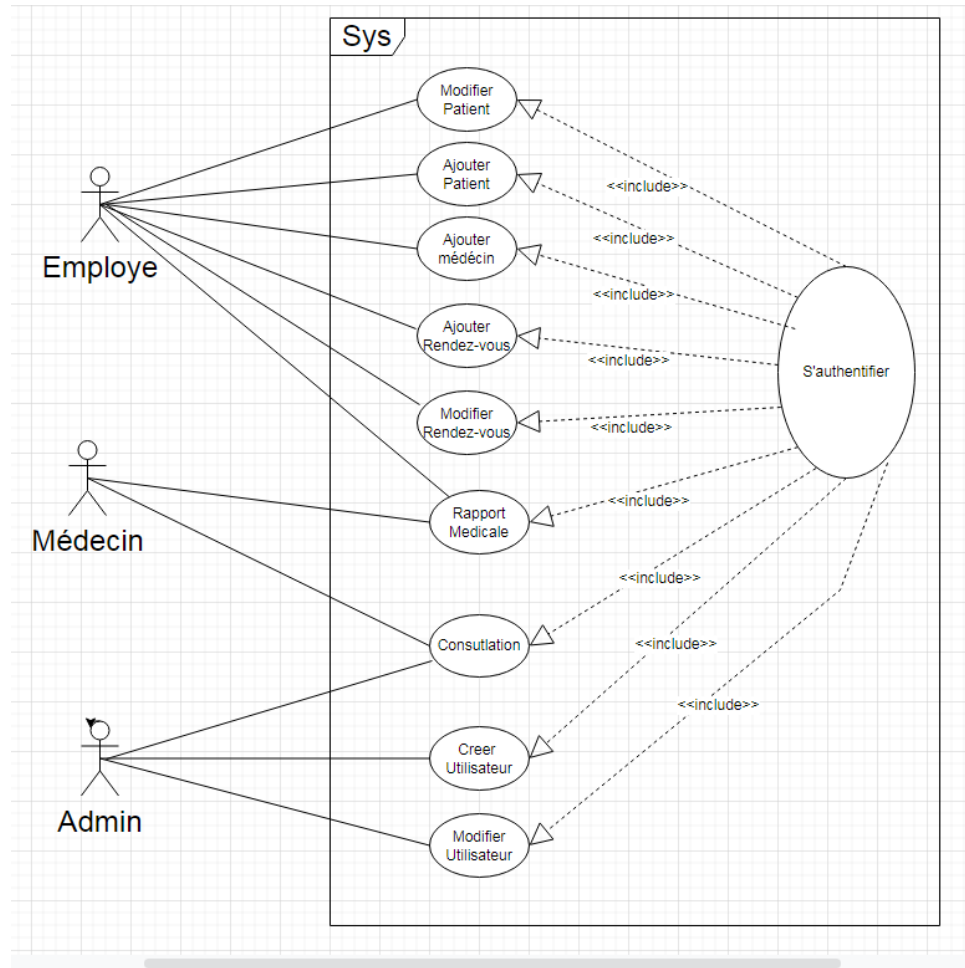


Figure 12 : Diagramme cas d'utilisation

1. **Employer :**
 - a. Créer patient et rendez-vous et les modifie.
 - b. Consulter les rendez-vous .
 - c. Upload les images médical .
2. **Médecin :**
 - a. Consulter rendez-vous .
 - b. Ecrire les rapports médicaux .
 - c. Uploadé les images médical.
3. **Administrateur :**
 - a. Créer les utilisateurs.
 - b. Modifier les rôles des utilisateurs.
 - c. Créer et modifier les patients

5.4. Diagramme de classe

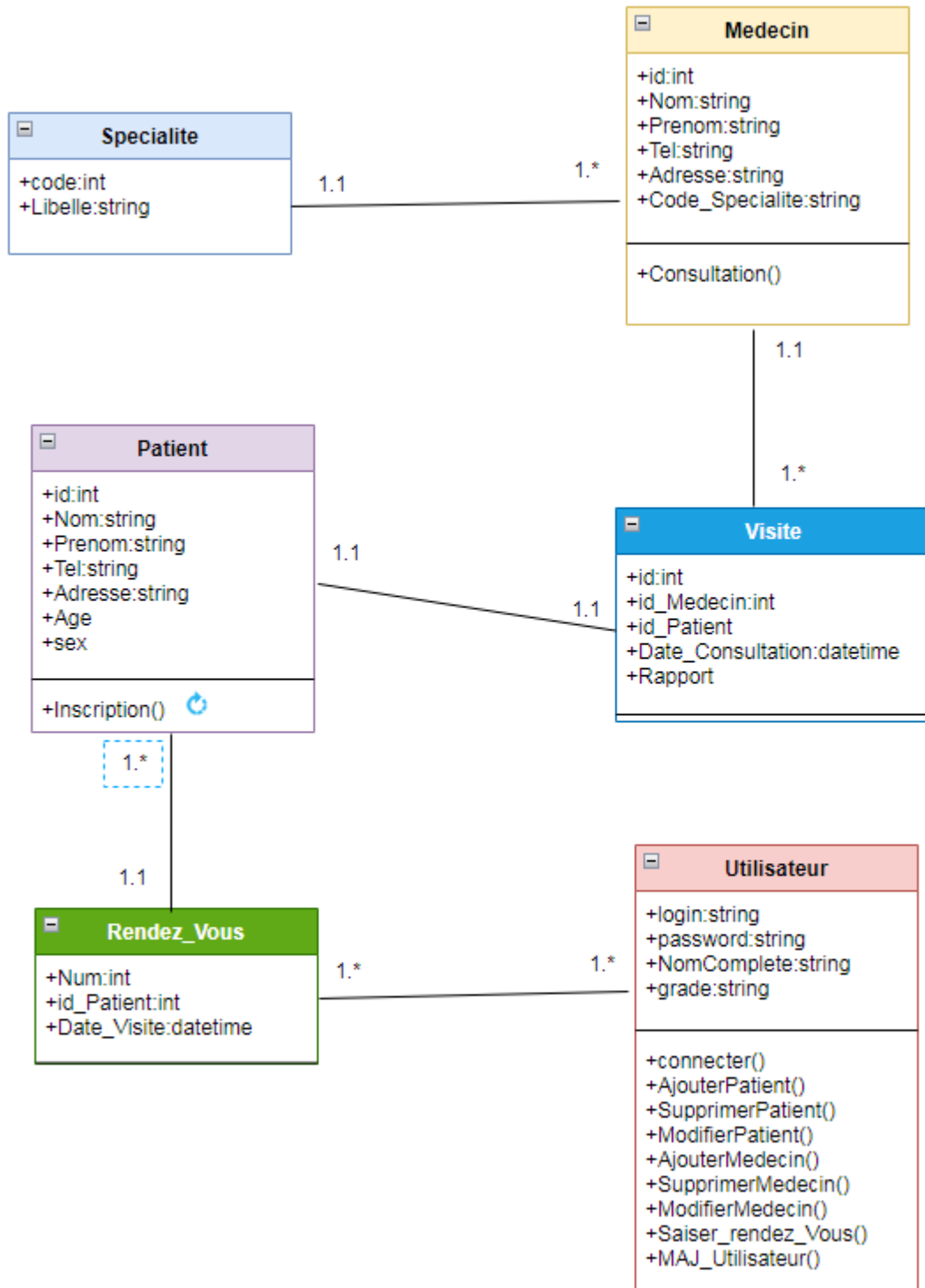


Figure 13 : Diagramme de classe (Qui présenter les classes d'application)

5.5. Diagramme de séquence

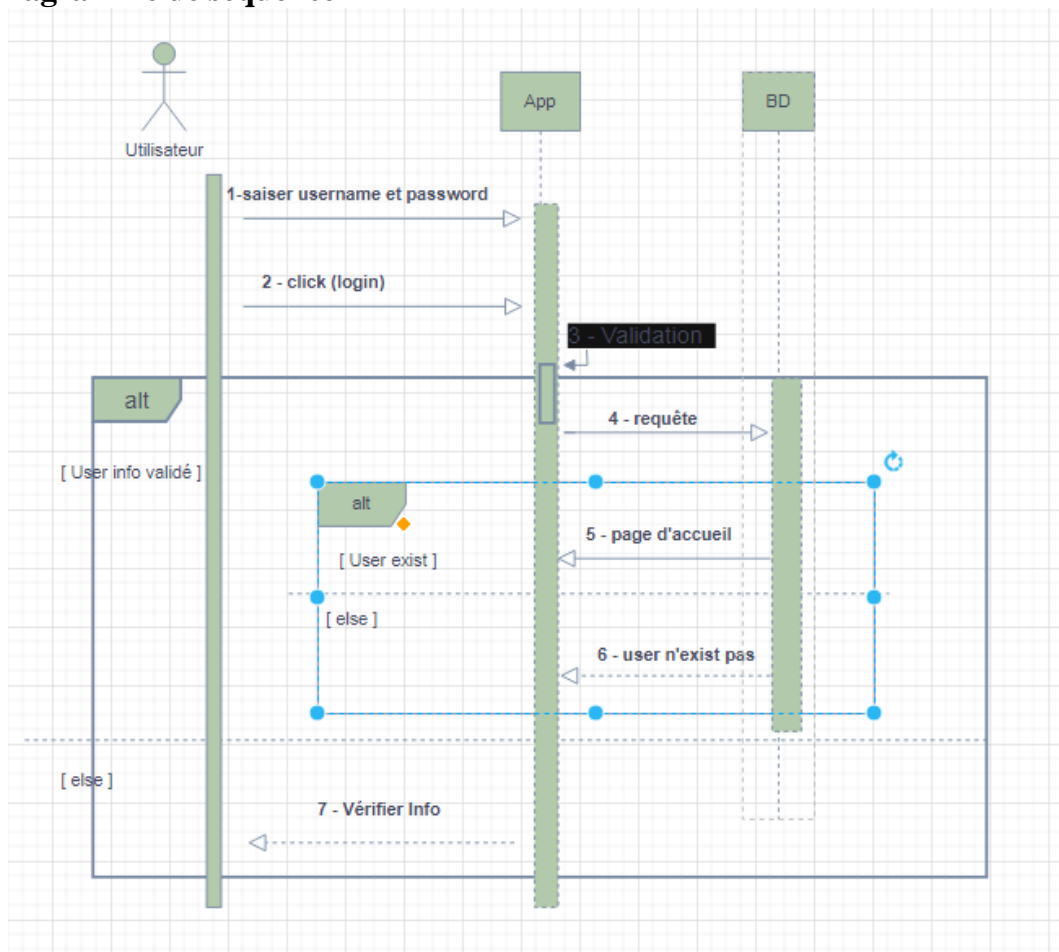


Figure 14: Diagramme de séquence (Authentification)

(Pour entrée saisis un nom et mot de passe)

- 1- Utilisateur saisir username et password puis click sur bouton login, l'application vérifier cet username et password si sont valide donc l'application chercher ce utilisateur dans la base de données si existe donc une nouvel fenêtre apparaitre (page d'accueil) .

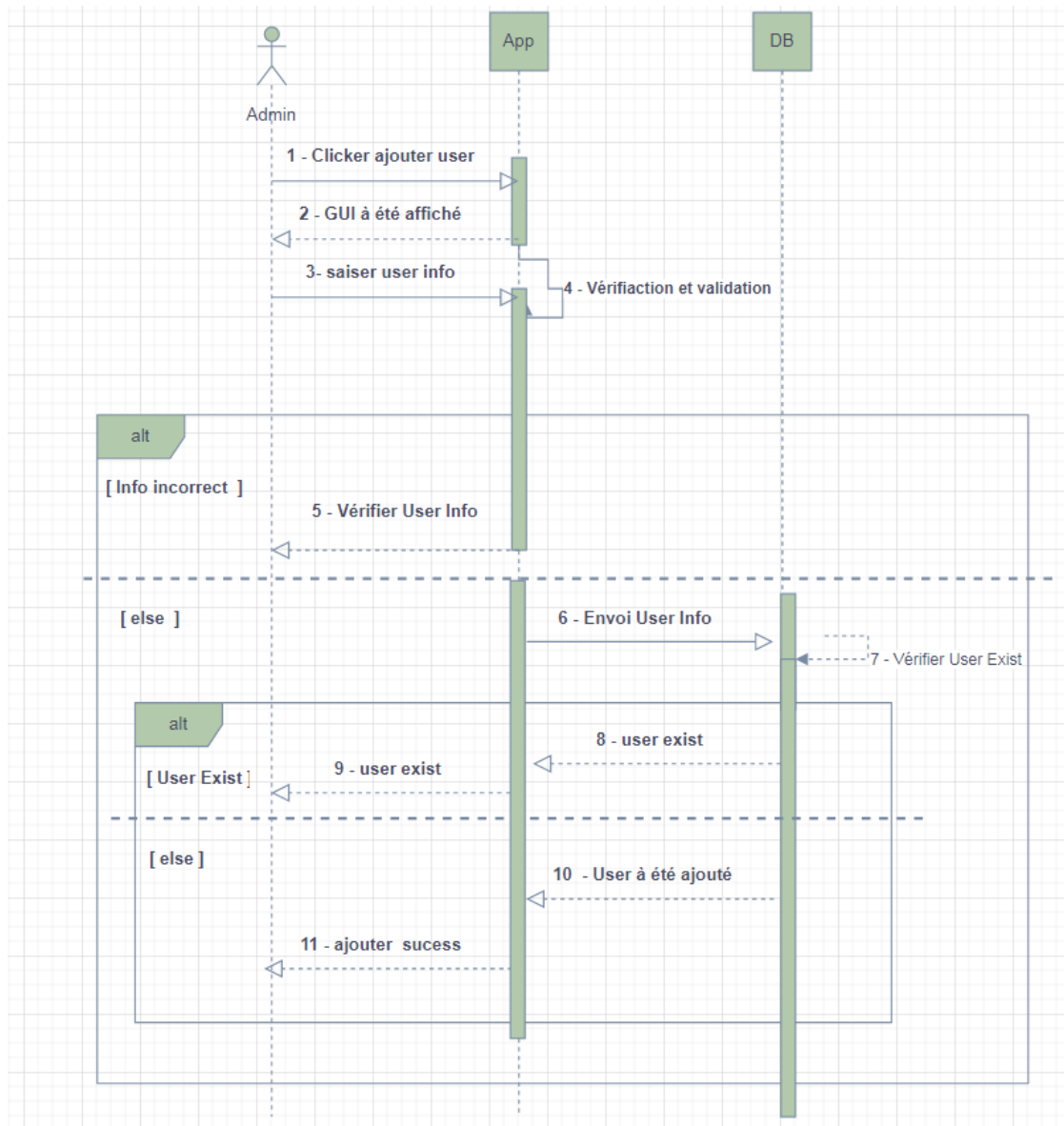


Figure 15: Diagramme de séquence (Création d'utilisateur)

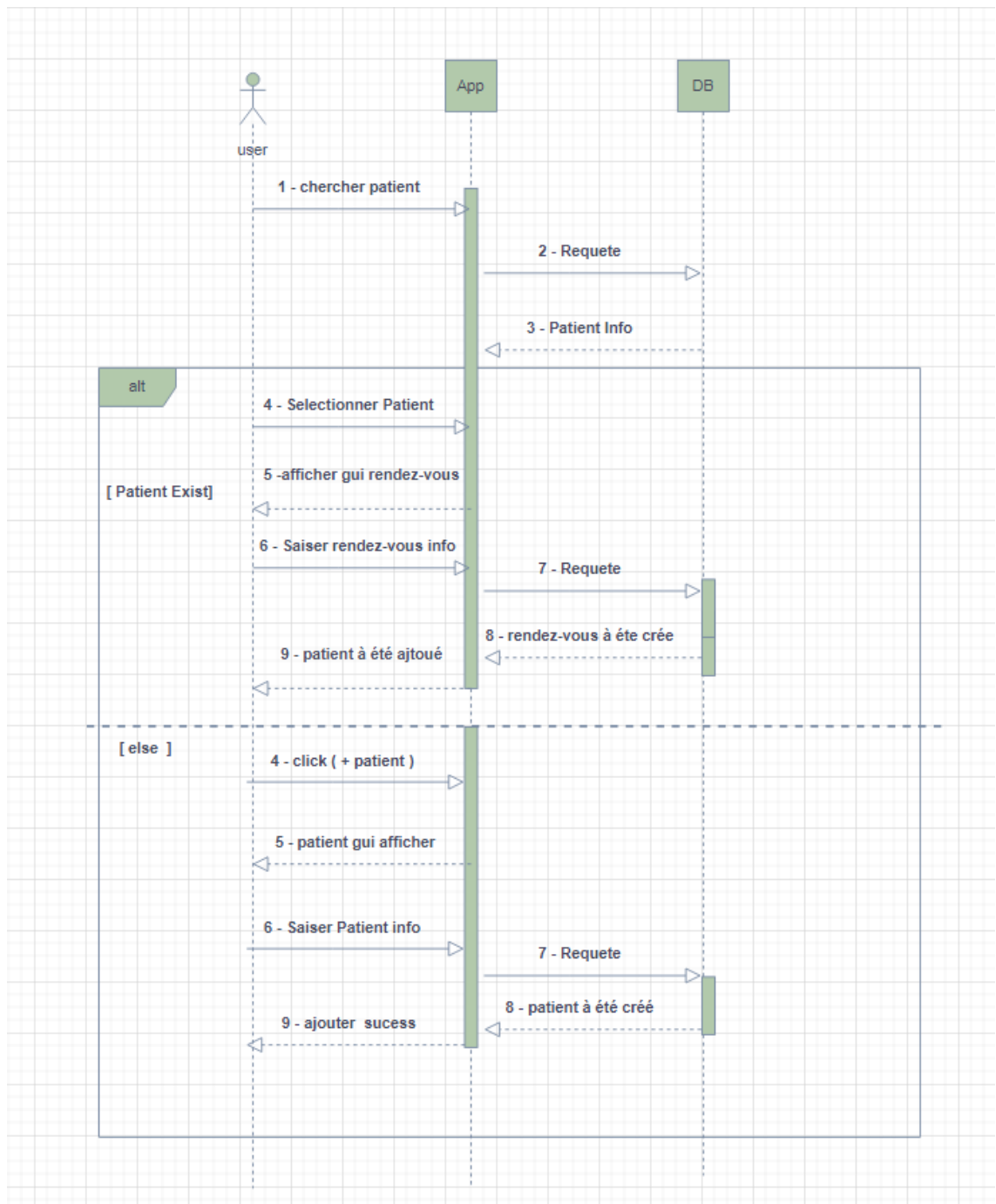
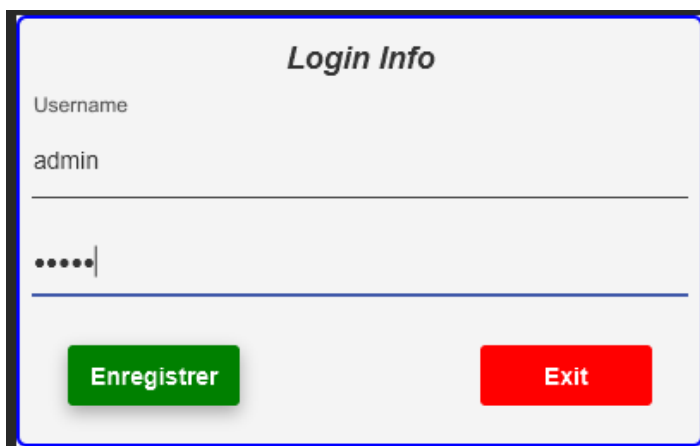


Figure 16 : diagramme de séquence (création de rendez-vous)

6. l'Application



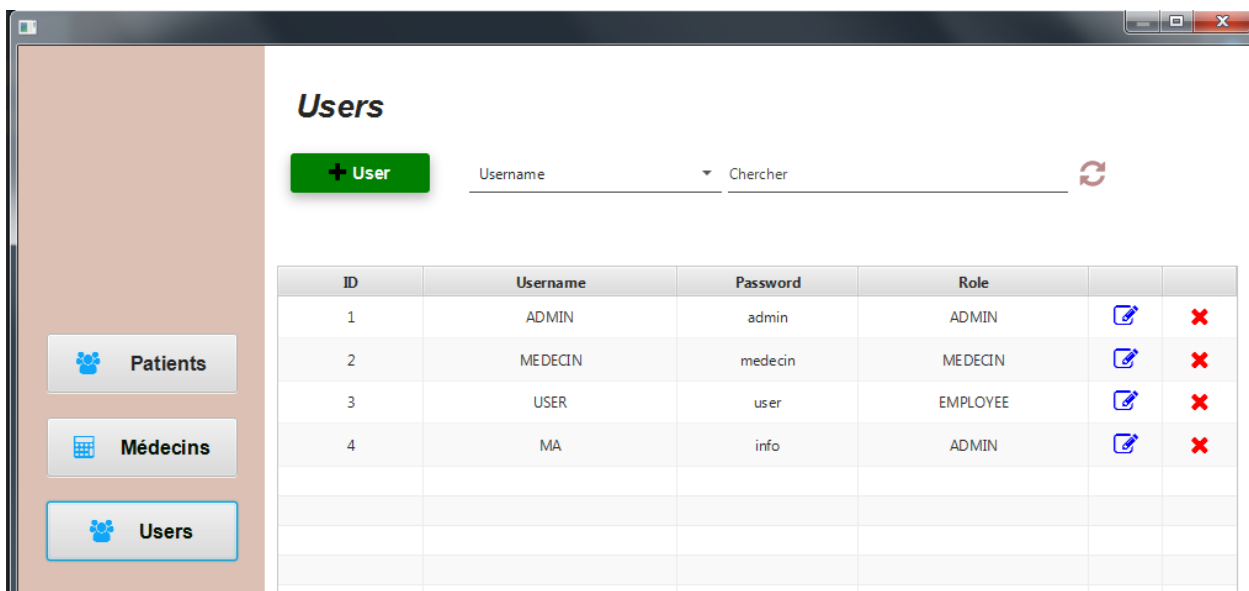
Username
admin

.....

Enregistrer Exit

Figure 17 : Interface entrée d'application par administrateur

(pour entrée nous utilisons nom "admin" et mot de passe "admin")



Users

+ User Username Chercher

ID	Username	Password	Role		
1	ADMIN	admin	ADMIN		
2	MEDECIN	medecin	MEDECIN		
3	USER	user	EMPLOYEE		
4	MA	info	ADMIN		

Figure 18 : Interface après entrée l'administrateur

Crée nouveau médecin, Employée et administrateur

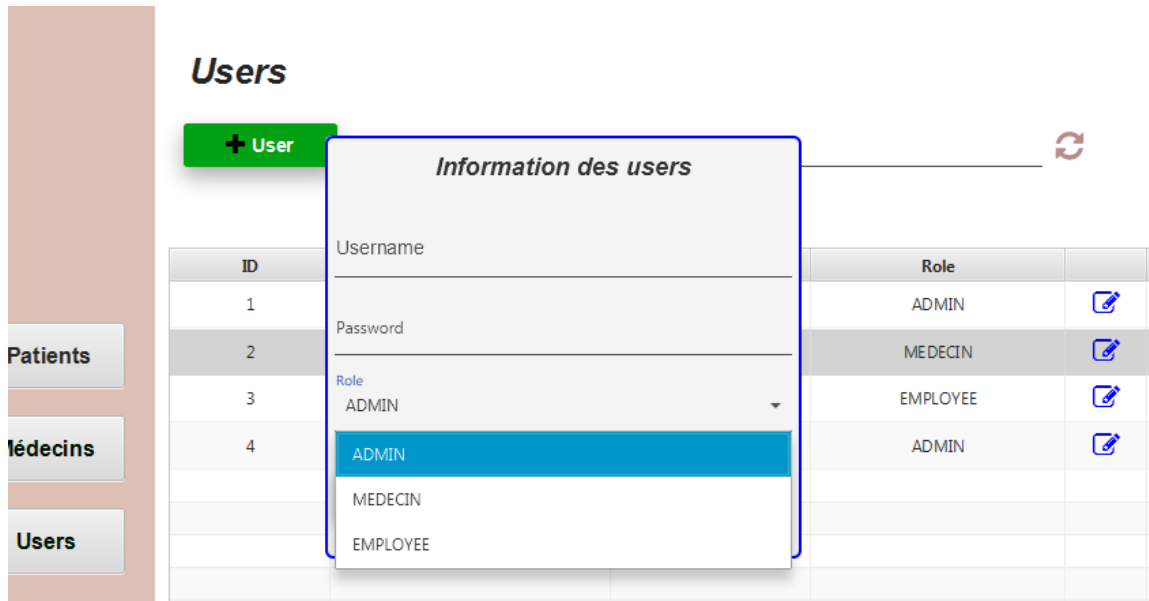


Figure 19: Interface pour crée nouveau médecin, Employée et administrateur

Patients

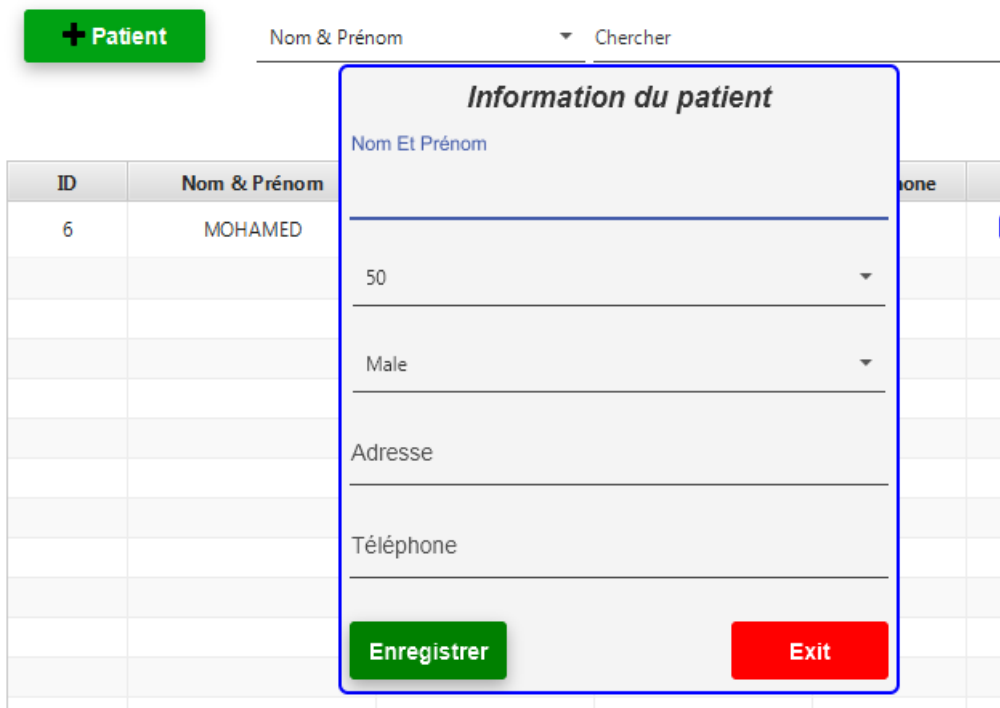


Figure 20: Interface pour entée un nouveau patient (malade)

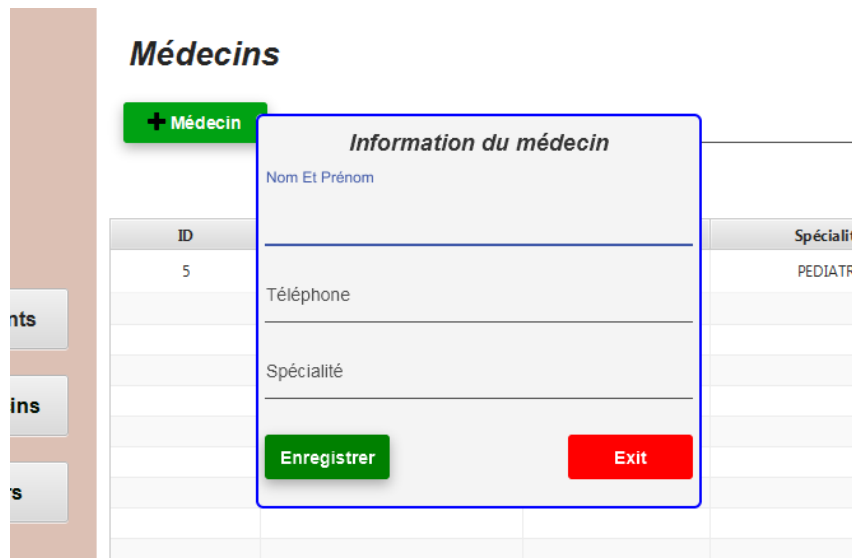


Figure 21: Interface crée médecin

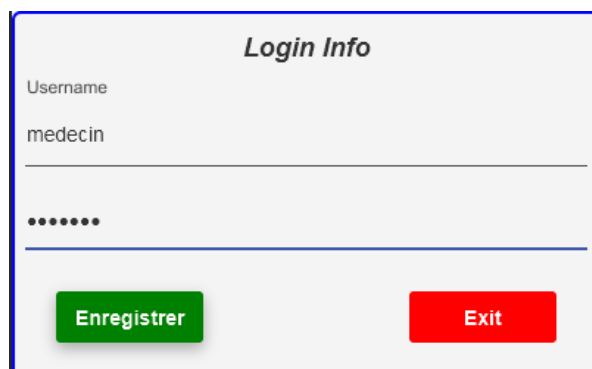


Figure 22: Interface entée par un médecin

(pour entée nous utilmisation nom "medecin" et mot de passe "medecin")

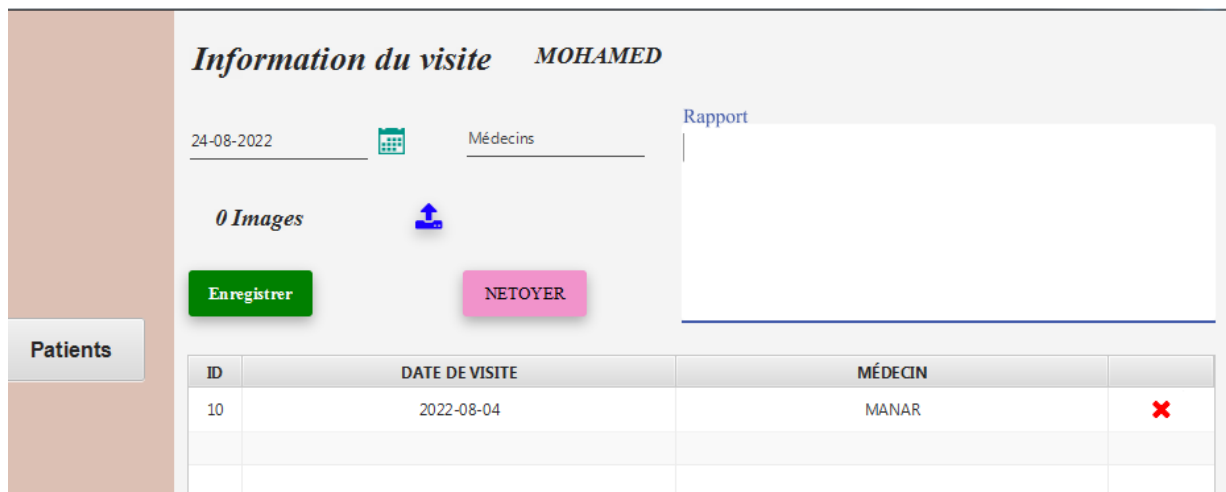


Figure 23: Interface autorisé au médecin

(Entée des images médicales de patient et le rapport médical)

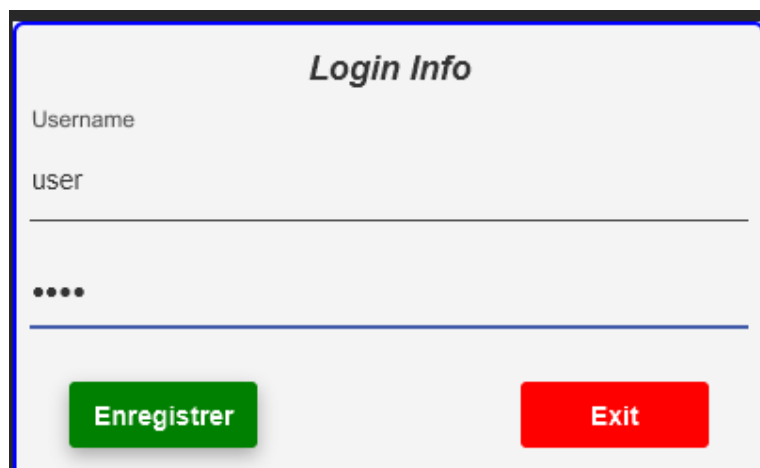


Figure 24: Interface entrée par un utilisateur

(pour entée : nom ‘user’ et mot de passe ‘user’)

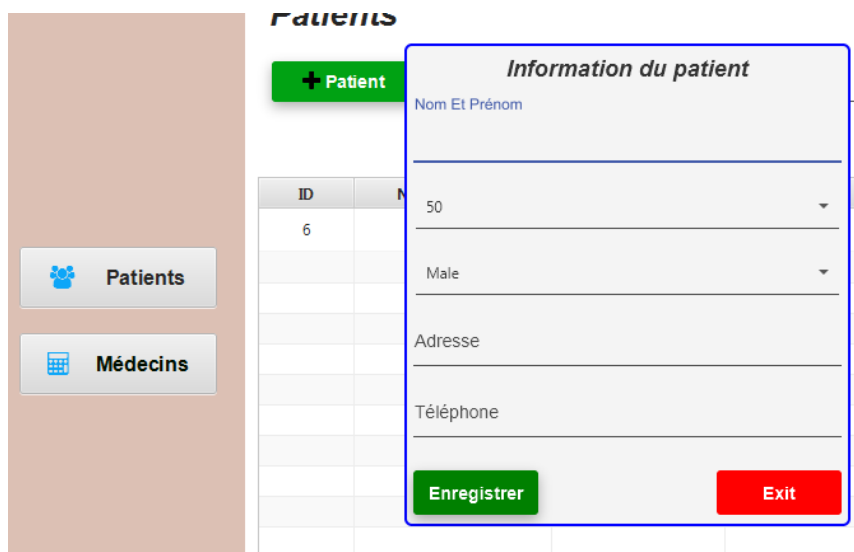


Figure 25: Entée un nouveau patient par l'employeur

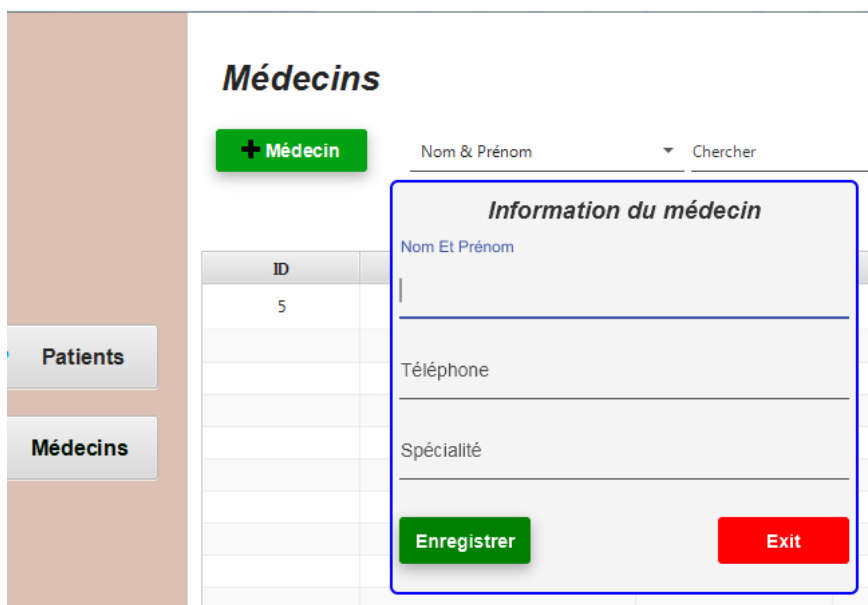


Figure 26: Créé un nouveau médecin par l'employeur

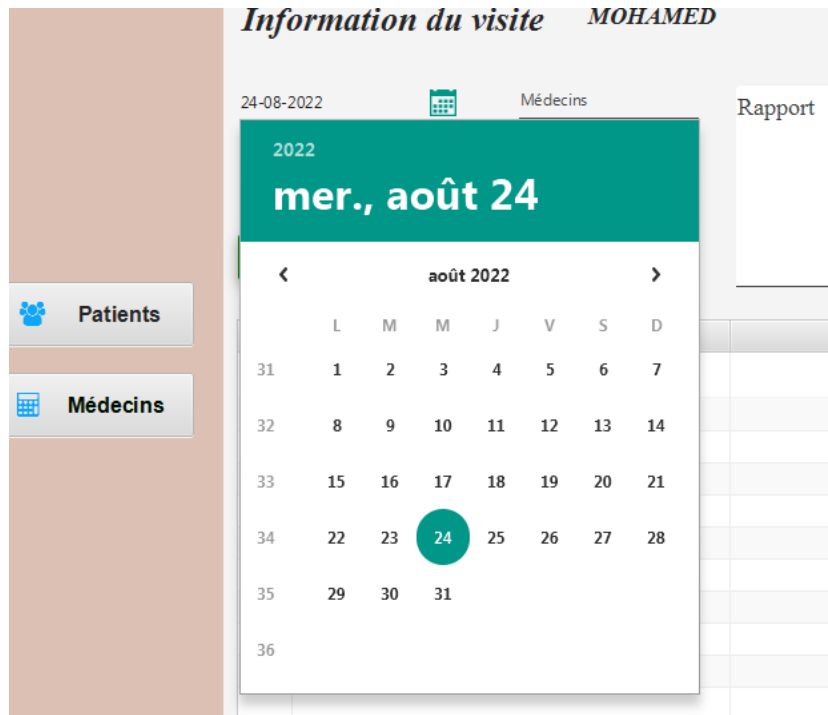


Figure 27: Crée un rendez-vous par un employeur

Télécharger Bibliothèque hadoop dans notre App''

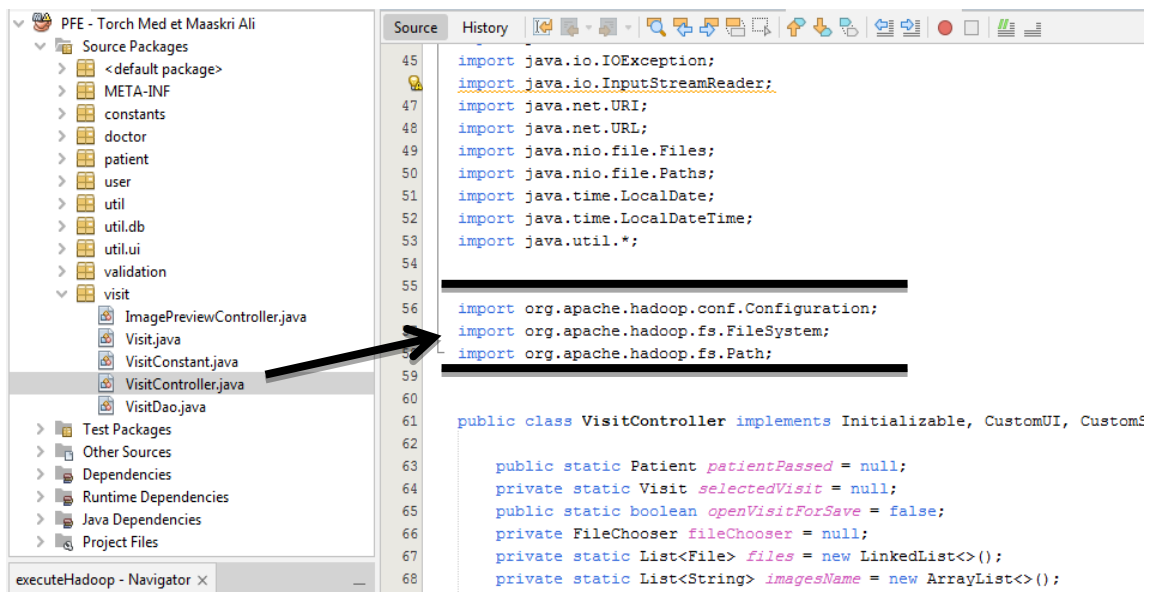


Figure 24 : Bibliothèque hadoop

Code pour télécharger les images médicales dans hadoop A chaque fois l'employeur télécharger un ou plusieurs images dans notre App

```
try {  
    System.out.println( x: "TorchMaaskri");  
  
    String HdfsUrl="hdfs://hadoop-master:9000";  
    Configuration conf=new Configuration();  
    conf.set( name: "fs.defaultFS", value: "hdfs://hadoop-master:9000");  
    conf.set( name: "fs.hdfs.impl", value: "org.apache.hadoop.hdfs.DistributedFileSystem");  
    FileSystem fs= FileSystem.get( uri: URI.create( str: HdfsUrl), conf);  
  
    String tp="/MedicalImages";  
  
    for(int i=0;i<files.size();i++){  
        fs.copyFromLocalFile( new Path( pathString: files.get( index: i)), new Path( pathString: tp));  
    }  
} catch (IOException ex) {  
    System.out.println( x: ex.getMessage());  
}  
}
```

Figure 25 : code java

Conclusion

Dans ce chapitre nous avons utilisé une simple application pour gérer un petit cabinet médical composé par plusieurs médecins et plusieurs employeurs.

Par cette application nous consultons les images médicales des patients et les rapports de chaque patient, et cette application peut télécharger les images médicales dans la plateforme hadoop en même temps.

Conclusion

générale

Serte que L'image médicale joue un rôle très important dans la santé contemporaine, comme outil et comme guide pour des procédures chirurgicales et thérapeutiques. Mais l'utilisation des images radiographies, ultrasonores, IRM, ... pose un grand problème de stockage et d'archivage. En plus du problème de stockage, si de telles images doivent être transmises via un réseau, la durée de la transmission est souvent trop longue. Pour pallier à tous ces problèmes.

Jusque-là, les travaux réalisés sur l'imagerie médicale ont surtout portés sur l'implémentation de banques de données d'images médicales et les différentes méthodes d'accès à ces informations. Notre approche est d'associer dans une même base de données des objets de type classique et des objets de type LOBs¹.

La finalité de ce travail est donc de mettre à la disposition du corps médical un outil de travail qui leur faciliterait la gestion des dossiers médicaux des malades en y incluant des données multimédias telles que les diagnostics de type texte classique et les images médicales (LOBs) qui peuvent être d'origine radiologique, IRM² ou scanner.

Bien que l'implémentation de notre base de données distribuée soit limitée à trois services médicaux (urgence, radiologie et exploration), les malades peuvent passer d'un service à un autre pour des subir les examens nécessaires sans déplacer leur dossier.

Les médecins traitants ont accès à toutes les informations concernant leur patient de façon transparente, sans se soucier du site où se trouve réellement l'enregistrement physique des données.

Si le travail réalisé dans cette première étape, permet de faciliter la gestion des données (consultation et stockage).

¹ - sont un ensemble de types de données conçus pour contenir de grandes quantités de données. Un LOB peut contenir jusqu'à une taille maximale allant de 8 téraoctets à 128 téraoctets selon la configuration de votre base de données. Le stockage des données dans des LOB vous permet d'accéder et de manipuler efficacement les données dans votre application

² - imagerie par résonance magnétique

Références

Conclusion générale

[1] Olivier BENDAVID, Biin The Cloud, Université Montpellier II, 18/06/2010.

[2] Big Data for Dummies, par Wiley Brand.

[3] Hadoop The Definitive Guide.3rd.Edition May 2012.Edition O'Reilly.

Thèse

[4] Mekideche Mounir, Conception et implémentation d'un moteur de recherche à base d'une architecture Hadoop (Big Data), Avril 2015.

[5] Annotation des images médicales dans un environnement parallèle et distribué 2020-2021.

PDF

[6] Benjamin Renault, Hadoop/Big Data, Université de Nice Sophia-Antipolis, 114p, 2013-2014.

[7] Bernard ESPINASSE, Introduction aux systèmes NoSQL (Not Only SQL), Ecole Polytechnique Universitaire de Marseille, 19p, Avril 2013.

[8] Charley CLAIRMONT, Introduction à HDFS Hadoop Distributed File System, HUG France SL2013, 20p, Mai 2013.

[9] Jonathan Lejeune, Hadoop : une plate-forme d'exécution de programme Map-reduce, École des Mines de Nantes, 83p, Janvier 2015.

[10] M.CORINUS, T.Derey, J.Marguerie, W.Techer, N.Vic, Rapport d'étude sur le Big Data, SRS Day, 54p, 2012.

Références Bibliographiques

[24] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, Piñeros M, et al. Observatoire mondial du cancer : « Cancer Today ». Lyon : Centre international de recherche sur le cancer, 2020 (<https://gco.iarc.fr/today>, consulté en juillet 2022).

[25] de Martel C, Georges D, Bray F, Ferlay J, Clifford GM. « Global burden of cancer attributable to infections in 2018: a worldwide incidence analysis ». The Lancet Global Health. 2020;8(2):e180-e190.

[26] Assessing national capacity for the prevention and control of non-communicable diseases: Report of the 2019 global survey. Genève : Organisation mondiale de la Santé, 2020.

[27] **Ruddon R. W. (2007)**. Cancer biology. Oxford. University Press.

[28] **Rouëssé J. (2011)**. Une histoire du cancer du sein en Occident: Enseignements et réflexions. Springer.

Conclusion générale

- [29] **Nogaret J. M. (2010).** Le cancer du sein: un regard optimiste vers l'avenir. Editions Academia.
- [30] **Flejou., J.F., (2012).** Généralité sur le cancer. Collège Français des Pathologistes (CoPath). France.
- [31] **Robert, J. (2013).** Biologie de la métastase. *Bulletin du cancer*, 100(4), 333-342.
- [32] **Coussy, F., Bonin, F., Azorin, P., Tariq, Z., Driouch, K. (2019).** Biologie des métastases et mécanismes moléculaires de leur formation. *Bulletin du Cancer*, 106(1), 24-36.
- [32] **IARC (International Agency for Research on Cancer). (2014).** World Cancer Report.
- [33] **Fouatih Z. A., Ammour N., Guetarn Ni., Bahloul N. N., Benrabah L. (2019).** Incidence du cancer à Oran selon le Registre du Cancer d'Oran Notifications pour l'année 2017 Cancer incidence in Oran according to the Oran Cancer Registry Notifications for 2017. *Revue des Sciences Médicales d'Oran*, 20.
- [34] **OMS (Organisation Mondial de la santé), (2003).** le cancer
[<https://www.who.int/fr/news-room/fact-sheets/detail/cancer>]
- [35] **OMS(Organisation Mondial de la santé), (2021).** le cancer
[<https://www.who.int/fr/news-room/fact-sheets/detail/cancer>]
- [36] **Alberts B., Bray D., Hopkin K., Johnson A., Lewis J., Raff M., Robert K., Walter P. (2012).** L'essentiel de la biologie cellulaire. Médecine sciences publications-Lavoisier.
- [37] **Hanahan D., Weinberg R. A. (2011).** Hallmarks of cancer: the next generation. *cell*, 144(5), 646-674.
- [38] **Hontaas A., (2014).** *Prise en charge des patients cancéreux à l'officine.* Th. Doct. Université Toulouse III
- [39] **Trouvé C. (2018).** Les effets indésirables cutanés, cutanéomuqueux et des phanères des anticancéreux: prise en charge officinale en dermo-cosmétique et autres thérapies complémentaires. Th. Doct. Université Toulouse III,
- [40] **Lecarpentier J., (2012).** Étude des facteurs modificateurs du risque de cancer du sein des femmes à risque génétique élevé. Th. Doct. Paris 11.
- [41] **Musgrove., E. A., Sutherland, R. L. (2009).** Biological determinants of endocrine resistance in breast cancer. *Nature Reviews Cancer*, 9(9), 631-643.
- [42] **Iwatsuki, M., Mimori, K., Yokobori, T., Ishi, H., Beppu, T., Nakamori, S., Mori, M. (2010).** Epithelial-

Conclusion générale

[43] **Ndoye M., Niang L., Gandaho K. I., Jalloh M., Labou I., Gueye S. (2014).** Cancer avancé de la prostate au

[44] **HAS (Haute Autorité de Santé) Française. (2012).** Rapport d'orientation - Cancer de la prostate. [http://www.has-sante.fr/portail/upload/docs/application/pdf/2012-04/rapport_dorientation_-

[45] **Balu-Maestro C., Chapellier C., Carrier P., Darcourt, ., Ettore F.,Raoust,I. (2005).** Imagerie dans le bilan

[46] **Ouedraogo V G. (2003).** Les cancers broncho-pulmonaires au centre hospitalier national Yalgado Ouedraogo :

[47] **Frusch N., Bosquee L., Louis R. (2007).** Le cancer du poumon. Epidémiologie et facteurs étiologiques. *Revue*

[48] **El jai A. (2009).** *Les cancers différenciés de la thyroïde à propos de 35 cas.* Th. Doct.

[49] **Bouklikha C, Sofiane D.(2014).**Cancer de la thyroïde. Université Aboubakr Belkaid Tlemcen. Mémoire doctorat

[50] **INC (Institut national du cancer). (2020).** « Cancer de la thyroïde : du diagnostic au suivi » en ligne, 19p.

[https://www.ecancer.fr/content/download/291475/4150776/file/Cancer_de_la_thyroide_du_d_iagnostic_au_suivi

[51] **Kechida S., A. F. A., et Teï A. (2017).** Etude générale sur la prévalence du cancer dans la région d'El-Oued.

[52] **Atbib Y. (2009).** Evaluation des taux des marqueurs tumoraux chez les hémodialyses chroniques (étude transversale a l'HMIMV-rabat).Th. Doct. 91p.

[53] **Lahaye L. (2018).** Mémoire, y compris stage professionnalisant [BR]-Séminaires méthodologiques intégratifs [BR]-Mémoire: Prescription d'examen d'imagerie médicale: enquête sur l'application de la prescription raisonnée.

[54] **Thomas, S. (2018).** Investigation de non-conformités microbiologiques d'endoscopes souples thermosensibles

[55] **Poirot C.,(2014).***L'information sur les effets indésirables de la chimiothérapie anticancéreuse: les besoins du patient et la place du pharmacien.* Th. Doct. Université de Lorraine .

[56] **Fernini K., et Soualah K., (2019).** Prévalence de cancer du sein dans la région d'AIN DEFLA.Université Djilali

Conclusion générale

[57] **Guennouni N. (2014).** Cancer colorectal et utilité clinique des marqueurs biologiques. universite Mohammed V. Th. Doct. 136p.

[58] **El aachab M. (2016).** Les Thérapies ciblées et leur place en onco-hématologie. Th. Doct.

[61] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, Ahmedin Jemal. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries, CA CANCER J CLIN 2018;68:394– 424.

Sites internet

[11] <http://www.journaldunet.com/solutions/analytics/big-data>, consulté le 06/06/2022.

[12] http://webcache.googleusercontent.com/search?q=cache:2DfXDe1Z_mEJ:www.aubay.com/fileadmin/user_upload/Publications_FR/Regard_Aubay__Big_Data_Web.pdf+&cd=1&hl=ar&ct=clnk&gl=dz, consulté le 2/03/2022.

[13] http://webcache.googleusercontent.com/search?q=cache:GTE9JIAU6mAJ:www.bigdataparis.com/guide/Guide_du_Big_Data_2013_2014.pdf+&cd=1&hl=ar&ct=clnk&gl=dz, consulté le 3/03/2022.

[14] <http://blog.ippon.fr/2013/05/14/big-data-la-jungle-des-differentes-distributionsopen-source-hadoop/>, consulté le 15/04/2022.

[15] <http://mbaron.developpez.com/tutoriels/bigdata/hadoop/introduction-hdfs-mapreduce/>, consulté le 27/04/2022 .

[16] <http://www.redsen-consulting.com/2013/06/big-data/>, consulté le 9/06/2022.

[17] http://fr.wikipedia.org/wiki/Entrep%C3%B4t_de_donn%C3%A9es, consulté le 9/04/2022.

[18] <http://www.technologies-ebusiness.com/langages/les-bases-no-sql>, consulté e. 6/06/2022

[19] <http://fr.wikipedia.org/wiki/Hadoop>, consulté le 10/04/2022.

[20] <http://blog.ippon.fr/2013/05/14/big-data-la-jungle-des-differentes-distributionsopen-source-hadoop/>, consulté le 10/04/2022.

[21] <http://www.devx.com/opensource/exploring-the-hadoop-distributed-file-systemhdfs.html>, consulté le 25/04/2022.

[22] <https://doc.ubuntu-fr.org>. consulté le 20/04/2022

[23] <https://gandalsmart.com>. consulté le 20/04/2022

[24] <http://www.speciation.net/Database/Instruments/Shimadzu-Europe/UV1650PC-;i1296> , visité le : 26/06/2022 à 07:59

Conclusion générale

[25] <https://www.jasco.de/en/content/V-630/~nm.13~nc.407/V-630-UV-VIS-Spectrophotometer.html> visité le : 25/06/2022 à 09:19

[60] <https://mse.utoronto.ca/curlook-materials-characterization-lab/x-ray-diffractometer-rigaku-miniflex-600/> visité le : 24/06/2022 à 22:21.

[62] [https://fr.wikipedia.org/wiki/UML_\(informatique\)](https://fr.wikipedia.org/wiki/UML_(informatique)) visité : 16/09/2022