

Livable 1

PROJET CLOUD HEALTHCARE UNIT

Oussama ALI-BEY - Alen AMIRBEKYAN

Adam MAHRAOUI - Antonin RUDONI



CESI 
ÉCOLE D'INGÉNIEURS

Table des matières

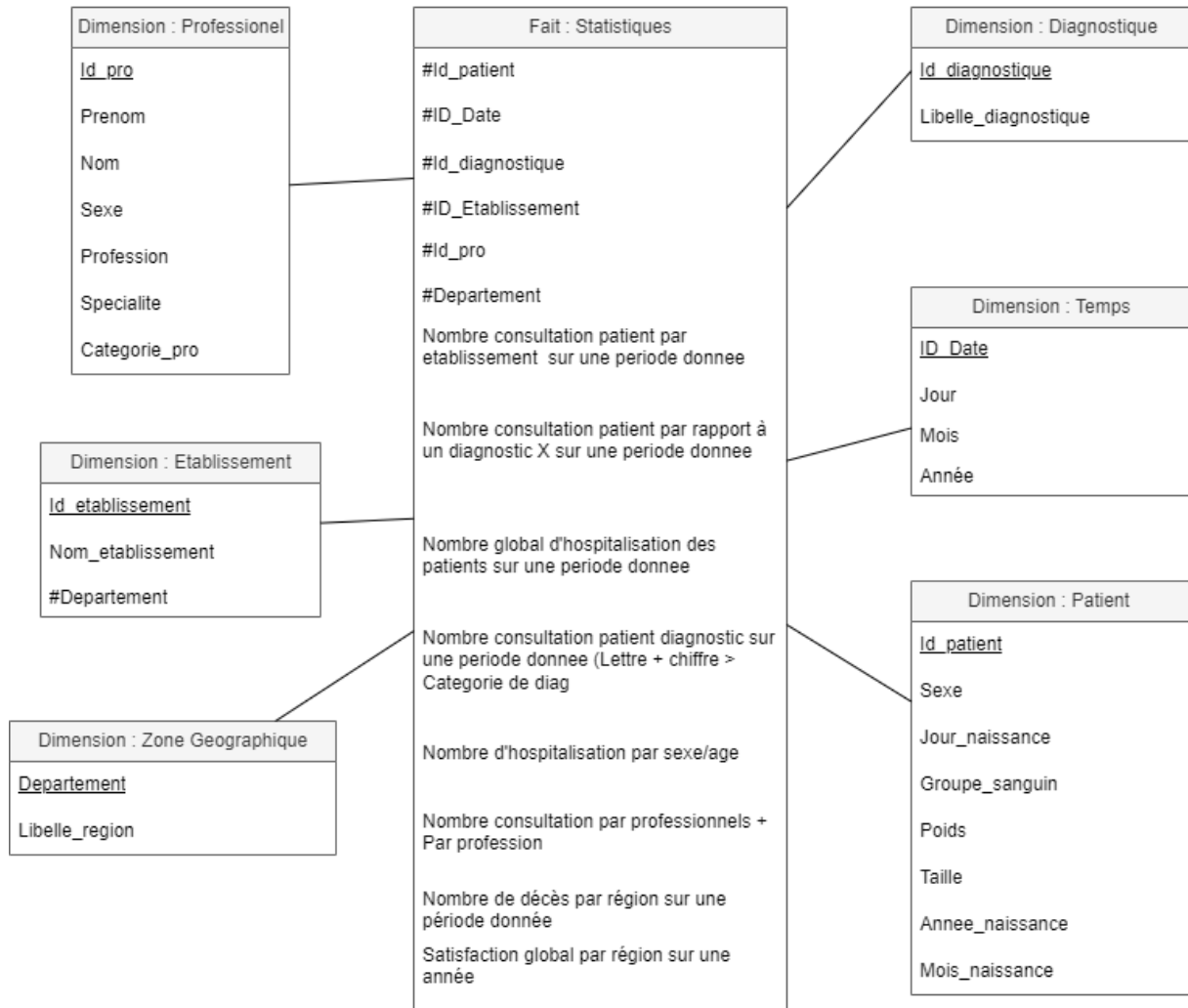
1.	Introduction	1
2.	Modélisation des différents axes d'analyse ainsi que les mesures	2
a)	Modification des demandes clients	3
b)	Dimension Professionnel	5
c)	Dimension Etablissement	5
d)	Dimension Zone Géographique	5
e)	Dimension Diagnostique	6
f)	Dimension Temps	6
g)	Dimension Patient	6
3.	Description de l'architecture de l'entrepôt de données.....	7
4.	Conclusion.....	7

1. Introduction

Le domaine de la santé évolue à un rythme très rapide, et les institutions médicales sont confrontées à la nécessité de s'adapter à de nouvelles techniques plus digitales. Dans cette perspective, le groupe CHU (Cloud Healthcare Unit) se trouve à un moment critique, cherchant à établir un entrepôt de données efficace pour gérer le flux continu de données de ses systèmes de gestion des soins et des systèmes FTP. Le but de ce projet est de concevoir une solution intégrale qui permette l'extraction, le stockage, l'intégration, l'exploration et la visualisation des données médicales de manière optimale. En collaboration avec le groupe CHU, nous devons identifier et répondre aux besoins spécifiques des praticiens et des administrateurs d'établissements, en leur fournissant des recommandations sur les outils d'intégration, de stockage et de visualisation les plus adaptés. Nos solutions vont permettre de contribuer à améliorer l'efficacité des services de santé en exploitant intelligemment leurs données.

2. Modélisation des différents axes d'analyse ainsi que les mesures

Schéma mutlidimensionnel



a) Modification des demandes clients

Pour satisfaire au mieux les besoins du client, nous avons décidé de préciser certains types d'analyses :

- Taux de consultation/hospitalisation des patients par rapport à un diagnostic X sur une période de temps Y.

Pour rendre les données pertinentes, nous avons pris la décision de catégoriser les diagnostics en nous basant sur le code CIM-10, en utilisant la catégorisation du tableau suivant :

I	A00– B99	Certaines maladies infectieuses et parasitaires
II	C00– D48	Tumeurs
III	D50– D89	Maladies du sang et des organes hématopoïétiques et certains troubles du système immunitaire
IV	E00– E90	Maladies endocriniennes, nutritionnelles et métaboliques
V	F00– F99	Troubles mentaux et du comportement
VI	G00– G99	Maladies du système nerveux
VII	H00– H59	Maladies de l'œil et de ses annexes
VIII	H60– H95	Maladies de l'oreille et de l'apophyse mastoïde
IX	I00–I99	Maladies de l'appareil circulatoire
X	J00–J99	Maladies de l'appareil respiratoire
XI	K00– K93	Maladies de l'appareil digestif
XII	L00– L99	Maladies de la peau et du tissu cellulaire sous-cutané
XIII	M00– M99	Maladies du système ostéoarticulaire, des muscles et du tissu conjonctif
XIV	N00–	Maladies de l'appareil génito-urinaire

	N99	
XV	O00– O99	Grossesse, accouchement et puerpéralité
XVI	P00– P96	Certaines affections dont l'origine se situe dans la période périnatale
XVII	Q00– Q99	Malformations congénitales et anomalies chromosomiques
XVIII	R00– R99	Symptômes, signes et résultats anormaux d'examens cliniques et de laboratoire, non classés ailleurs
XIX	S00– T98	Lésions traumatiques, empoisonnements et certaines autres conséquences de causes externes
XX	V01– Y98	Causes externes de morbidité et de mortalité
XXI	Z00– Z99	Facteurs influant sur l'état de santé et motifs de recours aux services de santé
XXII	U00– U99	Codes d'utilisation particulière

- Taux de consultation par professionnel

Pour rendre la statistique plus pertinente, nous avons décidé de calculer le taux de consultation par profession ou par spécialité, mais aussi par professionnel en les regroupant dans leurs professions ou par spécialité. Voici les professions qui seront utilisées comme références :

- "Chiropracteur"
- "Psychotherapeute"
- "Assistant de service social"
- "Podo-Orthésiste"
- "Chirurgien-Dentiste"
- "Pedicure-Podologue"
- "Masseur-Kinesithérapeute"
- "Psychologue"
- "Psychomotricien"
- "Orthoptiste"
- "Sage-Femme"
- "Pharmacien"
- "Technicien de laboratoire médical"
- "Oculariste"
- "Manipulateur ERM"

- "Audioprothesiste"
- "Orthopediste-Orthésiste"
- "Orthoprothesiste"
- "Orthophoniste"
- "Dieteticien"
- "Infirmier"
- "Osteopathe"
- "Medecin"
- "Ergotherapeute"
- "Epithésiste"
- "Opticien-Lunetier"

b) Dimension Professionnel

La dimension professionnel va permettre de décrire chaque professionnel de santé en fonction de son nom, son prénom, son sexe, sa profession, sa spécialité et sa catégorie professionnelle. Cela va permettre de calculer :

- le taux de consultation par professionnel

Mais aussi de laisser une ouverture sur les futurs besoins des clients en permettant de calculer :

- La répartition H/F au sein des spécialité, profession....

c) Dimension Etablissement

La dimension établissement va permettre de décrire chaque établissement de santé en fonction de son nom, son numéro FINESS qui sera utilisé en tant qu'identifiant (c'est un identifiant unique et définitif d'une entité juridique ou d'un établissement de santé), et le département.

Cela va permettre de calculer :

- Taux de consultation des patients dans un établissement X sur une période de temps Y

Mais aussi de laisser une ouverture sur les futurs besoins des clients en permettant de calculer :

- Répartition H/F en fonction de chaque établissement
- Répartition des établissements par région

d) Dimension Zone Géographique

La dimension Zone Géographique va permettre de décrire une localisation en fonction de son département (sous forme de chiffre) et du nom de la région (nous utiliserons le nom des 12 dernières régions de 2016).

Cela va permettre de calculer :

- Nombre de décès par localisation (région) et sur l'année 2019
- Taux global de satisfaction par région sur l'année 2020

Mais aussi de laisser une ouverture sur les futurs besoins des clients en permettant de calculer :

- Répartition des spécialités en fonction des régions

e) Dimension Diagnostique

La dimension diagnostique va permettre de caractériser les diagnostics en fonction de leur code CIM-10 qui servira d'identifiant (c'est une classification statistique codant les maladies, signes, symptômes, circonstances sociales et causes externes de maladies ou de blessures publiée par l'Organisation mondiale de la santé), ainsi que leur libellé. (voir plus haut dans la partie a))

Cela va permettre de calculer :

- Taux de consultation des patients par rapport à un diagnostic X sur une période de temps Y
- Taux d'hospitalisation des patients par rapport à des diagnostics sur une période donnée

Mais aussi de laisser une ouverture sur les futurs besoins des clients en permettant de calculer :

- Taux de consultation des patients par rapport à un diagnostic X par rapport à une région
- Taux de consultation des patients par rapport à un diagnostic X par rapport à un groupe, sanguin, un IMC, un âge

f) Dimension Temps

La dimension Temps va permettre de décrire une temporalité (pour mesurer des périodes) en fonction du jour, du mois, de l'année et de l'identifiant de la date.

Cela va permettre de calculer :

- Taux de consultation des patients dans un établissement X sur une période de temps Y
- Taux de consultation des patients par rapport à un diagnostic X sur une période de temps Y
- Taux global d'hospitalisation des patients dans une période donnée Y
- Taux d'hospitalisation des patients par rapport à des diagnostics sur une période donnée

g) Dimension Patient

La dimension Patient va permettre de décrire chaque patient (de manière anonyme afin de respecter le RGPD) en fonction d'un identifiant, son sexe, sa date de naissance(jour, mois, année), son poids, son groupe sanguin et sa taille.

Cela va permettre de calculer :

- Taux de consultation des patients dans un établissement X sur une période de temps Y
- Taux de consultation des patients par rapport à un diagnostic X sur une période de temps Y
- Taux global d'hospitalisation des patients dans une période donnée Y

- Taux d'hospitalisation des patients par rapport à des diagnostics sur une période donnée
- Taux d'hospitalisation par sexe, par âge

Mais aussi de laisser une ouverture sur les futurs besoins des clients en permettant de calculer :

- Le taux consultation pour un diagnostic en fonction du groupe sanguin, du poids, du sexe

3. Description de l'architecture de l'entrepôt de données

Pour ce projet portant sur la BIGDATA, nous avons reçu une grande quantité de données à traiter et à analyser en vue de produire un tableau de bord pertinent pour nos clients. Ces données se présentent sous divers formats, mais la majorité d'entre elles sont en format CSV ou Excel. Nous avons donc opté pour l'utilisation d'un DATA WAREHOUSE pour stocker cette masse de données. L'avantage de cette solution de stockage réside dans sa capacité à centraliser et à structurer efficacement une grande variété de données, principalement issues de fichiers CSV ou Excel. En utilisant un entrepôt de données tel que HIVE, nous bénéficions de performances élevées et d'une évolutivité adaptée à nos besoins. Nous avons donc mis en place ce stockage à l'aide de HIVE.

4. Conclusion

Maintenant que nous avons une idée précise de la future structure de notre architecture BIG-DATA, il est temps de la mettre en place avec les outils adaptés et de la façonner selon nos besoins.