**Compte Rendu Projet Big-Data**

Réponses aux questionnaires :

1. Expliquez pourquoi a-t-on recours aux technologies Big Data dans la réalisation de ce type de système ?

Étant donné la nature du système nous optons pour ces technologies afin de faire usage des 5V proposés par cette architecture.

Volume : Dans le domaine de la santé et notamment pour le cas de notre travail ou plusieurs capteurs ou appareils IoT (Internet des objets) sont déployés, les volumes de données à collecter et analyser sont considérables et en augmentation constante.

Vitesse : De plus en plus souvent, les données doivent être collectées et traitées en temps réel, afin de faire des analyses prédictifs sur l’état de santé d’un patient.

Variété : Les données peuvent prendre des formes très variées et très hétérogènes en fonction de des différents capteurs installés

Véracité : La véracité ou fiabilité des données est notamment menacée par les diversités des points de collecte, par la multiplication des formats de données.

Valeur : Dans un contexte d'infobésité, il s'agit d'être capable de se concentrer sur les données ayant une réelle valeur et étant actionnables.

1. Quelle est la partie qui relève du mode ‘batch’ et quelle partie relève du mode ‘streaming’ ?

Batch : Les données transférées dans la base des données mongo pour consultation soit par les médecins, infirmiers ou chercheurs pour des fins d’études et autres.

Streaming : Les données des patients en état d’alerte transférées en temps réel après analyse et prédictions aux médecins ou infirmiers urgentistes afin de secourir ce patient.

1. Expliquez dans le contexte de ce système e-health le rôle de chaque couche et proposez, pour chaque couche, la technologie Big Data à utiliser en justifiant votre choix.

Nous avons plusieurs couches dans ce systèmes avec différents rôles à savoir :

Source des données : Les capteurs récupèrent les données des patients

Collection : Les Gateways permettent de collecter plusieurs données, alors à ce niveau nous proposons l’utilisation du Machine Learning pour faire de la classification afin de traiter les données avant l’envoi aux serveurs.

Traitement en mode ‘batch’ : On utilisera un cluster Kafka-Mongo créé sous docker afin de pouvoir utiliser les topics de ce dernier dans le stockage des données reçues des gateways et de là nous utiliserons des connecteurs mongo pour envoyer ces données dans la base de donnée.

Traitement en mode ‘streaming’ : Les consommateurs seront utilisés pour envoyer des alertes aux médecins et infirmiers urgentistes, en récupérant les données sur le topic dédié depuis le messages en queue sur le producteur

Stockage : La base de donnée mongo sera dédiée au stockage.

1. Le ‘Message Broker’ est un composant très utile dans ce type de système. Justifiez son utilité́ ?

Il permet de gérer la file d’attente des messages reçus du producteurs et s’assurer que le message a été supprimé après qu’il ait reçu un accusé de réception du destinateur lui signifiant qu’il a bien consommé le message, ainsi un message ne peut être lu qu’une et une seule fois et par un seul destinateur.

1. Quels sont les ‘topics’ qui vous paraissent pertinents pour ce système au niveau du ‘Message Broker’ ?

Nous voterons pour l’utilisation de deux topics dont :

Alerte : Pour gérer la partie ‘streaming’

Normal : Pour les données des clients en état normal.

Architecture du système

