

BIG DATA

VOLTRON

NDIANE ELABDOURAHIM DIOP

ADRIEN ADAMISTE

ARTHUR SIMONIAN

Introduction :

Pour répondre à l'appel d'offre, nous vous proposons de collecter, examiner, gérer, traiter et exploiter l'ensemble des données issues du mécanisme mis en place par l'expert IOT.

Les données collectées sont :

- La température
- L'humidité
- La pression
- La quantité de dioxyde de carbone

La concentration en microparticules par m3 contenu dans l'air dans les blocs opératoires de l'hôpital.

Nous allons mettre en place une base de données pour stocker et sécuriser ces données. Elles seront ensuite envoyées à l'équipe IA qui feront des études sur ces données.

Un tableau de bord va être mis en place pour la visualisation de ces données dans le but d'examiner les données.

Choix des solutions :

Nous allons faire l'étude de deux solutions de gestion des données à savoir Elasticsearch et MongoDB.

Elasticsearch et MongoDB sont les deux magasins de données distribués les plus populaires utilisés pour gérer les données NoSQL. Ces deux technologies sont hautement évolutives et reposent sur une conception orientée document. Il existe cependant des différences entre les deux technologies, et il est important de comprendre ces différences afin de choisir celle qui convient à notre cas d'utilisation.

1. Elasticsearch

Elasticsearch est un moteur de recherche et d'analyse distribué, gratuit et ouvert pour tout type de données, y compris les données textuelles, numériques, géospatiales, structurées et non-structurées. Elasticsearch a été conçu à partir d'Apache Lucene et a été lancé en 2010 par Elasticsearch N. V. (maintenant appelé Elastic). Réputé pour ses API REST simples, sa nature distribuée, sa vitesse et sa scalabilité, Elasticsearch est le composant principal de la Suite Elastic, un ensemble d'outils gratuits et ouverts d'ingestion de données, d'enrichissement, de stockage, d'analyse et de visualisation. Couramment appelée la Suite ELK (pour Elasticsearch, Logstash et Kibana), la Suite Elastic comprend désormais une riche collection d'agents de transfert légers, appelés les agents Beats, pour envoyer des données à Elasticsearch.

La vitesse et la scalabilité d'Elasticsearch, ainsi que sa capacité à indexer de nombreux types de contenus signifient qu'il peut être employé dans différents cas d'utilisation.

Kibana

Kibana est un greffon de visualisation de données pour Elasticsearch publié sous la licence libre Apache version 2.

Il intègre des histogrammes en temps réel, des graphes linéaires, des camemberts et des cartes. Kibana intègre également des applications comme Canvas, qui permet aux utilisateurs de créer des infographies dynamiques personnalisées basées sur leurs données, et sur Elastic Maps pour visualiser des données géospatiales.

2. MongoDB

MongoDB est une base de données orientée document écrite en C++ avec une conception à l'esprit pour gérer des téraoctets de données répartis sur plusieurs géolocalisations. Dans MongoDB, vous pouvez créer plusieurs bases de données et chaque base de données peut avoir plusieurs collections (tables). MongoDB est également une base de données sans schéma qui prend en charge des fonctionnalités de sécurité intégrées telles que l'authentification, le contrôle d'accès et le cryptage.

MongoDB charts

MongoDB Charts est un outil pour créer des représentations visuelles de vos données MongoDB. La visualisation des données est un élément clé pour fournir une compréhension claire de vos données, mettre en évidence les corrélations entre les variables et faciliter le discernement des modèles et des tendances au sein de votre ensemble de données. MongoDB Charts simplifie la communication de vos données en fournissant des outils intégrés pour partager et collaborer facilement sur les visualisations.

Modèle de données

Nom	Type
Date	Date Time
CO2	Double
Humidité	Double
Pm10	Double
Pm25	Double
Température	Double
VOC	Double

VOC : Ce sont des polluants précurseurs de l'ozone, et certains d'entre eux sont considérés comme cancérogènes pour l'homme.

PM10 désigne les particules dont le diamètre est inférieur à 10 micromètres (noté μm , $1\ \mu\text{m} = 10^{-6}\text{ m}$ c'est-à-dire 1 millionième de mètre ou encore 1 millième de millimètre). Le diamètre des particules fines PM2.5 et PM1 sont inférieurs respectivement à 2.5 et à 1 μm .

Tarification

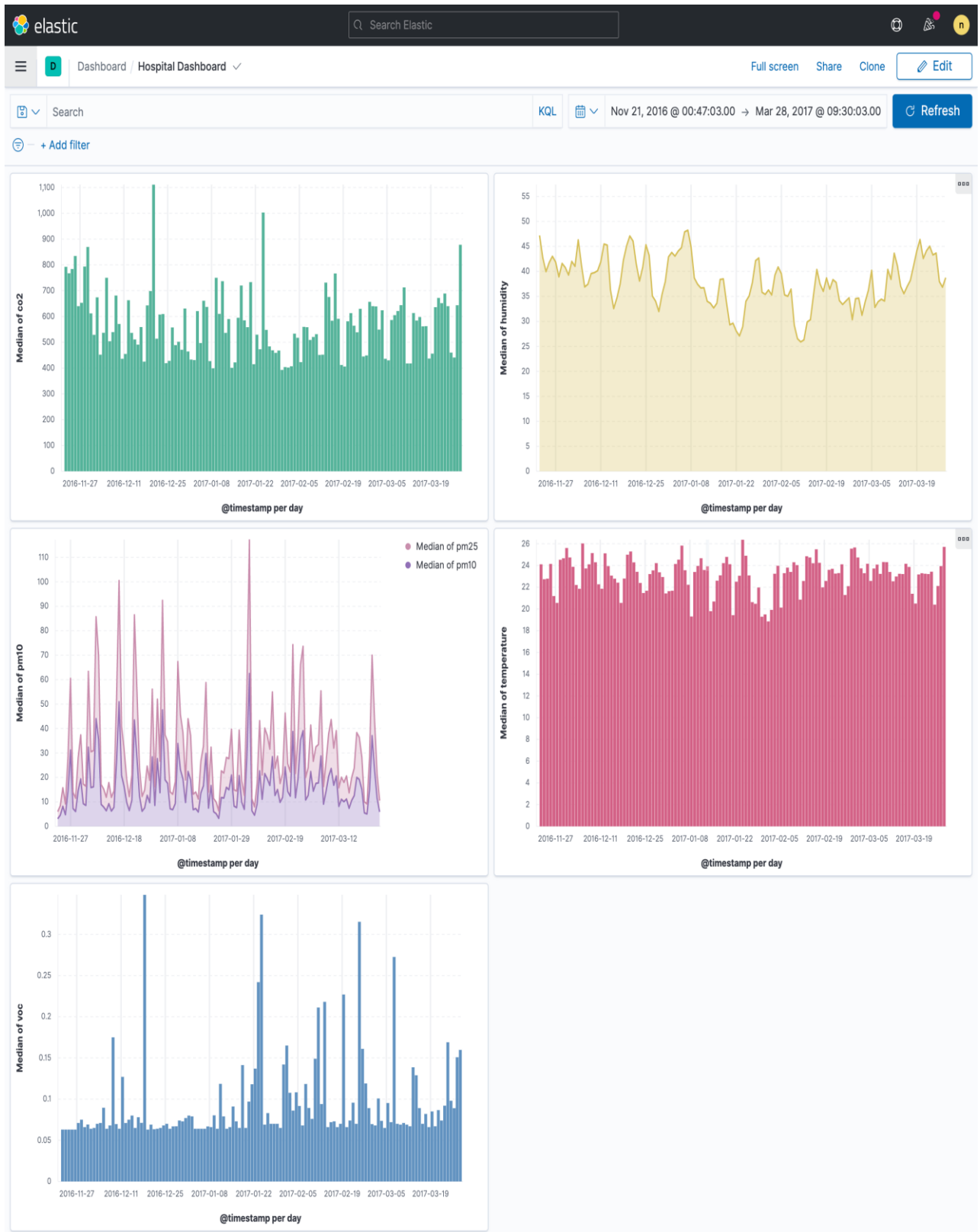
Kibana:

- Standard (16 USD/mois)
- Gold (19 USD/mois)
- Platinum (22 USD/mois)
- Entreprise (30 USD/mois)

MongoDB Charts

- Shared (\$0/mois) limité
- Dedicated (\$57/mois)
- Multi-region (\$95/mois)

POC avec Kibana



POC avec MongoDB Charts

