
REMERCIEMENTS

Je tiens à remercier toutes les personnes qui ont contribué à la réussite de mon stage et de près ou de loin à la rédaction de mon rapport.

Tout d'abord, je tiens à commencer par Mme Manel Boumaiza mon encadrante à Sagemcom pour son accueil, le temps qu'elle a consacré pour moi et son partage de connaissances. La réalisation des tâches n'aurait pas pu s'accomplir sans son aide précieuse.

Je remercie aussi vivement mon encadrante Esprit Mme Ines Slimen. Ses conseils académiques et sa disponibilité ont été un atout pour moi pour me situer durant la réalisation et la rédaction du rapport.

J'adresse également mes remerciements à toutes les personnes de Sagemcom sans qui je n'aurais pas pu profiter d'une telle expérience. Je nommerai surtout M. Hichem Maatallah, Nader Gharsallaoui et toute l'équipe ETL.

Pour finir, je remercierai toute personne m'ayant fourni de l'aide durant, avant et après ce stage.

TABLE DES MATIÈRES

Introduction générale	1
Chapitre 1 : Contexte du projet	4
Introduction	5
1. Présentation de l'organisme d'accueil	5
2. Présentation du travail demandé	7
Conclusion.....	10
Chapitre 2 : État de l'art	11
Introduction	12
1. Internet des Objets	12
2. Solution Sagemcom pour l'Internet des Objets	15
3. Missions du stage	17
Conclusion.....	19
Chapitre 3 : Architecture de la solution proposée.....	20
Introduction	21
1. Introduction au « Big Data »	21
2. Étude et choix techniques	25
3. Entrepôt de donnée Druid.....	35
4. Indexation en temps réel.....	37
5. Architecture détaillée du système	39
Conclusion.....	41
Chapitre 4 : Identification des indicateurs de performance.....	42
Introduction	43

1. Les indicateurs de performance	43
2. Les tableaux de bord	46
Conclusion.....	48
Chapitre 5 : Réalisation	49
Introduction	50
1. Environnement de travail	50
2. Déploiement du système d'analyse.....	51
3. Implémentation de l'application de visualisation	52
4. Chronogramme.....	71
Conclusion.....	72
Conclusion et perspectives	73
Webographie	75

LISTE DES FIGURES

Figure 1 : Logo Sagemcom	5
Figure 2 : Organigramme Sagemcom Software & Technologies.....	6
Figure 3 : Le modèle Incrémental	10
Figure 4 : Secteurs d'activités de l'Internet des objets	14
Figure 5 : Les couches logique de l'infrastructure IoT	14
Figure 6 : Architecture réseau de la solution IoT de Sagemcom	15
Figure 7 : Membres de l'alliance LoRa	16
Figure 8 : Couverture géographique du réseau LoRa	16
Figure 9 : Architecture d'une solution Big Data	23
Figure 10 : Architecture globale du système.....	27
Figure 11 : Architecture de la distribution Hortonworks	29
Figure 12 : Architecture de la distribution Cloudera	30
Figure 13 : Architecture de la distribution Map-R	31
Figure 14 : Architecture globale d'un cluster Druid.....	37
Figure 15 : Architecture détaillé de la solution de supervision de la qualité de service	40
Figure 16 : Architecture du déploiement de Hortonworks.....	51
Figure 17 : Architecture de déploiement du cluster Druid.....	52
Figure 18 : Interface d'authentification utilisateur	52
Figure 19 : Interface dashboard administrateur partie 1	55
Figure 20 : Interface dashboard administrateur partie 2	56
Figure 21 : Interface de la liste des objets	57
Figure 22 : Interface tableau de bord objet	58

Figure 23 : Interface de la liste des applications	59
Figure 24 : Interface tableau de bord application	60
Figure 25 : Interface de la liste des passerelles	60
Figure 26 : Interface tableau de bord passerelle	61
Figure 27 : Interface de la liste des réseaux	62
Figure 28 : Interface tableau de bord réseau partie 1.....	62
Figure 29 : Interface tableau de bord réseau partie 2.....	63
Figure 30 : Interface de l'activité réseau.....	64
Figure 31 : Interface dashboard administrateur réseau partie 1	66
Figure 32 : Interface dashboard administrateur réseau partie 2	67
Figure 33 : Interface dashboard « Administrateur d'une zone géographique ».....	68
Figure 34 : Interface dashboard « Administrateur application »	69
Figure 35 : Interface dashboard du client	70

LISTE DES TABLEAUX

Tableau 1 : Domaines d'application de l'IoT	13
Tableau 2 : Tableau comparatif des technologies Big Data.....	27
Tableau 3 : Tableau comparatif des distributions Hadoop.....	31
Tableau 4 : Tableau de définition des objectifs d'analyse	47
Tableau 5 : Chronogramme de la planification des tâches du projet.....	71

INTRODUCTION GÉNÉRALE

L'Internet des Objets est devenu un réel phénomène en 2014 et 2015. Des nouveaux réseaux de télécommunication "machine to machine" dédiés à la communication entre objets connectés et les infrastructures internet existantes ont vu le jour. Parmi ces réseaux nous pouvons distinguer entre les réseaux sans fils à faible portée et les réseaux sans fils de longue distance.

Tous les acteurs de télécommunications sont donc appelés à jouer un rôle dans la mise en place de ces nouvelles technologies. Sagemcom, le leader européen sur le marché des terminaux communicants, propose déjà une offre intégrée et de bout en bout pour l'Internet des Objets basée sur le protocole LoRaWANTM pour les réseaux étendus de longue portée.

Avec l'infrastructure de communication, chaque opérateur de réseau IoT doit fournir à ses clients une visibilité claire sur les performances réseau et le niveau de qualité des connexions cellulaires à travers des solutions de mesure de qualité de service. La qualité de service est l'un des concepts les plus importants dans les réseaux des objets connectés car ces réseaux sont exploités souvent par des applications critiques.

Sagemcom cherche à mettre en place une solution de supervision en temps réel de la qualité de service des objets connectés à son réseau LoRa. Cette solution présente l'état du réseau et les indicateurs de performance clés spécifiques à ce réseau dans des tableaux de bord.

C'est dans ce cadre que s'inclue ce projet de fin d'étude, qui consiste à mettre en place un système de supervision de la qualité de service d'un réseau IoT. Ce système se base sur les techniques d'analyse de données et les outils de gestion du volume massif de données. Les résultats des analyses sont présentés dans une application web.

Ce présent rapport s'organise sur cinq chapitres principaux.

Le premier chapitre « Contexte du projet » consiste à présenter le cadre général du projet. Ce chapitre contient une présentation de l'organisme d'accueil, une idée globale sur le sujet ainsi qu'une étude de la partie existante et la solution proposée.

Le deuxième chapitre « État de l'art » est consacré à la compréhension de l'Internet des objets, la présentation de la solution offerte par Sagemcom ainsi que la définition des objectifs à atteindre.

Le troisième chapitre « Architecture de la solution proposée » détaille l'architecture du système déployé et les études des choix faites sur chaque module de la solution.

Le quatrième chapitre « Identification des indicateurs de performance » présente les différents indicateurs de performance dégagés pour la mesure de la qualité de service du réseau.

Le cinquième chapitre « Réalisation » présente le travail effectué pour la réalisation de la solution. Il contient une présentation des activités d'implémentation et comporte également la planification du projet et les principales interfaces graphiques.

L'ensemble sera clôturé par une conclusion et une ouverture de perspectifs concernant ce stage.

CHAPITRE 1 :

CONTEXTE DU PROJET

Introduction

Dans ce chapitre, nous exposons le contexte général du projet. Nous présentons en premier lieu l'organisme d'accueil, puis le travail demandé et nous terminons par l'explication de la méthodologie de travail que nous avons suivi.

1. Présentation de l'organisme d'accueil

Ce projet est élaboré dans les locaux de Sagemcom Software & Technologies ou « SST » qui est une filiale de recherche et développement de la multinationale Sagemcom. Nous présentons tout d'abord l'entreprise Sagemcom. Puis, nous détaillons l'activité de « SST »

1.1. Sagemcom

Sagemcom est un groupe français leader européen sur le marché des terminaux communicants à haute valeur ajoutée (décodeurs, box Internet, compteurs électriques,...). En 2015, Le Groupe réalise 1,26 milliard d'euros de chiffre d'affaires, compte 4200 salariés présents dans plus de 40 pays. Il opère sur plusieurs secteurs et principalement dans :

- « Broadband » (décodeurs, box TV/Internet),
- « Smart City » (Compteurs intelligents),
- « Internet of Things » (Objets connectés). ^[1]



Figure 1 : Logo Sagemcom

Plus de 22 millions de terminaux sont conçus, fabriqués et livrés chaque année dans le monde entier par Sagemcom. Il est détenu à 30% par ses salariés et à 70% par Carlyle, une société de gestion d'actifs mondiaux. Le groupe est piloté par une équipe de direction stable, dont les membres sont présents au sein du groupe depuis plus de 20 ans, et à la tête de Sagemcom depuis sa sortie du groupe Safran en 2008. ^[1]

Le siège social de Sagemcom est situé en France à Rueil-Malmaison (92), centre névralgique du groupe où se trouve également un centre de recherche et développement (R&D).

[1]

Sagemcom a pour ambition de devenir l'un des leaders mondiaux des terminaux communicants à forte valeur ajoutée en adressant des marchés dynamiques. Pour cela, il vise à être le premier à proposer à ses clients des produits customisés intégrants les dernières ruptures technologiques.^[1]

1.2. Sagemcom Software & Technologies

1.2.1. Présentation

Sagemcom Software et Technologies abrégé SST est un centre de recherche et de développement, filiale du groupe Sagemcom, qui a été créé le 18 juillet 2005. Sa mission est de concevoir et développer des produits et des logiciels pour l'ensemble des activités de Sagemcom.

La direction de SST est structurée en trois pôles d'activités : ATR (Activités Terminaux Résidentiels), ADT (Activités Décodeurs de Télévision numérique), AET (Activités Énergie et Télécom), et des équipes support :

- Le service RH,
- Le service Informatique,
- Le service Services Généraux et Environnement,
- Le service Qualité,
- Le service Entité Technologie Logicielle.

L'organigramme de SST est présenté dans figure 2. Ce projet a été élaboré au sein de l'équipe « Anticipation » qui fait partie du service Entité Technologie Logicielle « ETL ».^[2]

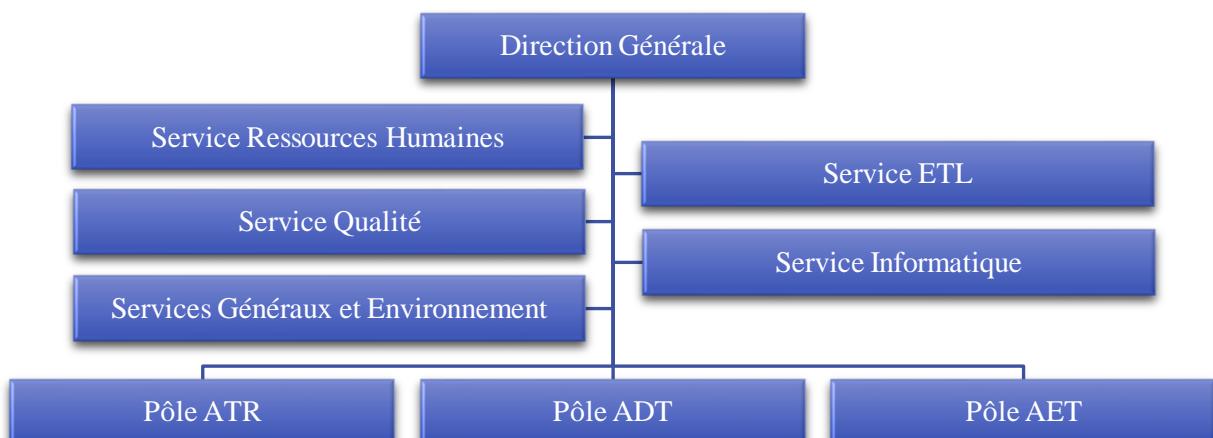


Figure 2 : Organigramme Sagemcom Software & Technologies

1.2.2. Équipe Anticipation

L'équipe Anticipation est une activité d'innovation à Sagemcom. Cette équipe est répartie entre la filiale SST en Tunisie et le siège social à Rueil-Malmaison. Sa mission est d'étudier de nouvelles technologies et de nouveaux protocoles afin de garder les produits existants en phase avec les avancements technologiques ou pour explorer des nouveaux marchés. L'activité Anticipation prépare des preuves de concept (POCs) et des démonstrations pour anticiper les besoins des clients.

2. Présentation du travail demandé

2.1. Objectifs du projet

Pour les solutions offertes dans le cadre de son secteur d'activité dédié à l'Internet des objets, Sagemcom reste fidèle à son métier qui consiste à mettre sur le marché des produits communicants à forte valeur ajoutée. Il propose une offre intégrée et de bout en bout basée sur le standard ouvert LoRaWAN™ (Long Range Wide -area ou réseau étendu de longue portée). Cette offre s'étend sur des modules radio destinés aux Objets connectés, jusqu'à l'ensemble du réseau d'infrastructure, stations de bases, cœur de réseau, ou encore solutions logicielles de redistribution des données aux opérateurs de service.

LoRa est un réseau composé de plusieurs objets communicants à fortes contraintes (calcul, énergie, stockage, etc.). Chaque objet génère des données (température locale, taux d'humidité, etc.) qu'il doit ensuite transmettre à un serveur de collecte via des passerelles dédiées à ce réseau. Dans ce réseau, les liens sans fil sont à faible débit et à taux de pertes élevés.

Le réseau LoRa est mis au service d'applications qui nécessitent des opérations complexes citons à titre d'exemples les soins de santé et la surveillance industrielle. Cette utilisation soulève des nouveaux problèmes liés aux exigences de qualité de service (QoS). Ainsi, le fait de supporter la QoS s'avère difficile en présence des ressources très limitées des objets connectés et de la faible qualité des liens radios ainsi que des environnements de déploiement.

Pour cela, Sagemcom cherche à créer un système de supervision en temps réel des indicateurs de performance du réseau qui va permettre de mesurer la QoS du réseau. Le point de départ de notre application est l'ensemble des données générées par les objets qui sont collectées au niveau de serveur réseau LoRa.

2.2. Solution existante

Nous avons trouvé, au démarrage de notre projet, une solution de collecte et de persistance de données déjà en place. Une architecture basée sur une base de données NoSQL pour stocker les données brutes générées par les objets connectés LoRa. Notre travail était de construire une solution d'analyse de ces données et de reporter les résultats dans des tableaux de bord.

Cependant, nous avons relevé plusieurs limites pour cette architecture en relation, principalement, avec nos besoins de supervision et d'analyse en temps réel, une latence entre l'envoie des données par les objets et leur disponibilité à l'analyse. Et un problème d'optimisation pour nos types d'analyses qui sont généralement des requêtes d'agrégation.

2.3. Solution proposée

Notre mission dans ce projet est répartie en trois phases. En premier lieu, nous préparons un module de collecte, filtrage et nettoyage de données qui seront récupérées depuis les différentes sources de données client. Cette étape doit prendre en compte la diversité des messages reçus et la diversité des méthodes de réception de données.

La deuxième étape consiste à mettre en place un système de stockage et d'analyse des données reçues. Ce système sera responsable d'effectuer les analyses nécessaires au calcul des indicateurs de performance. En plus, il doit s'interfacer avec la solution de visualisation qui sera réalisée durant la troisième phase.

Finalement nous créerons les tableaux de bord de visualisation des indicateurs de performance. Ces dashboards doivent être spécifiques au métier du consommateur ainsi que la mise en valeur des graphes de visualisation les plus intéressants.

2.4. Processus de gestion du projet

Dans ce type de projet, le modèle classique du cycle en V est déconseillé. Il faut au contraire construire la solution progressivement, en prévoyant quelques itérations comprenant des interactions avec les futurs utilisateurs. Dans ce type de projet, les données sont au cœur des attentes. Par conséquent, un dialogue permanent avec les utilisateurs des données doit être établi.

La méthode de gestion adoptée pour notre projet se base sur le processus de réalisation par incrément. Cette méthodologie permet une réalisation moins complexe, une intégration plus

aisée et la création des livrables en continue même si la solution ne soit pas finalisée. Ce modèle convient aux projets de grande envergure. La figure 3 représente le modèle incrémental et les étapes que nous avons déjà décrits.

Nous commençons tout d'abord notre projet par une étude approfondie du protocole réseau LoRaWAN pour se familiariser avec son mode de fonctionnement, les données échangées et les facteurs clés dans la mesure de la QoS. Durant cette période nous mettrons la main sur la partie existante du projet.

À l'issue de cette étude nous fixerons le cadre du projet par la précision des résultats attendus. Nous allons aussi fixer les objectifs intermédiaires pour la réalisation de ce projet.

Le premier objectif intermédiaire consiste à définir les indicateurs de performance réseau spécifiques au réseau LoRa. Nous allons donc analyser les données échangées par les objets connectés, définir une liste d'indicateurs cibles et étudier l'importance de chaque indicateur. À l'issue de cet objectif, nous devons avoir une liste d'indicateurs pertinents qui est validée par les experts du réseau LoRa.

Quant à l'étape suivante nous allons étudier les technologies qui nous permettent d'analyser les messages des objets connectés, une architecture de système sera alors décidée. Après, Nous commençons le déploiement de l'architecture et l'intégration des données. Un processus de test et de validation va être lancer en parallèle avec la tâche de déploiement pour s'assurer que le système d'analyse soit conforme à la liste des objectifs fixés d'une part et fidèle à la liste des indicateurs de performance d'autre part.

La partie finale est consacrée au développement de l'outil de visualisation. Nous allons présenter les indicateurs de performance dans des graphes ciblés. L'ensemble de ces graphes sera servi dans des tableaux de bord spécifiés par utilisateur.

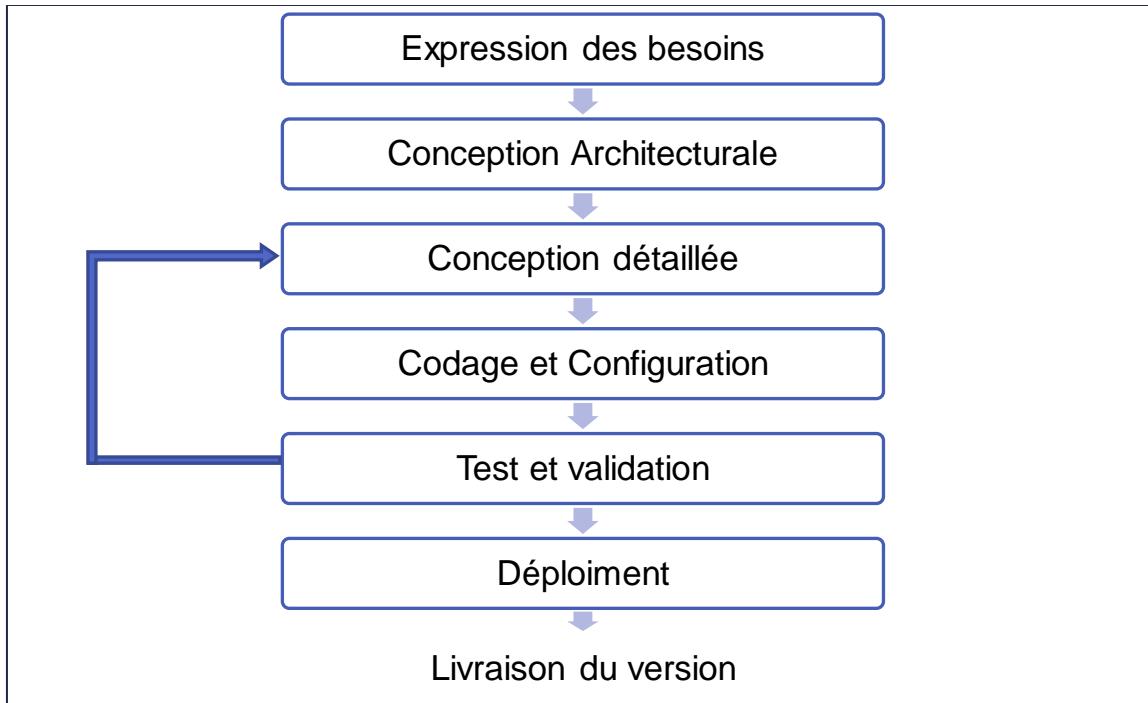


Figure 3 : Le modèle Incrémental

Conclusion

Après une présentation de l'organisme d'accueil et après un premier diagnostic des objectifs de ce projet, nous allons déceler les problématiques et la solution envisagée qui sera détaillée dans le prochain chapitre.

CHAPITRE 2 : ÉTAT DE L'ART

Introduction

Les objets connectés sont présents dans notre vie quotidienne et le seront de plus en plus, nous citons comme exemple : nos montres, notre électroménager, nos voitures, nos appareils de santé, etc. L'objectif de rendre tout objet communicant est de nous donner de plus en plus d'informations sur nos comportements, nos habitudes, afin de nous aider à mieux gérer notre quotidien ou notre santé. Nous consacrons ce chapitre à l'étude du monde d'Internet des objets, afin d'expliquer ce qu'il signifie et comment Sagemcom compte en devenir un acteur majeur.

1. Internet des Objets

1.1. Définition

L'Internet des Objets (en anglais « Internet of Things » ou « IoT ») désigne la connexion de ces objets à un réseau plus large, que ce soit directement (par Wi-Fi par exemple), par l'intermédiaire du smartphone de l'utilisateur (souvent via une connexion Bluetooth) ou grâce à des protocoles de communication qui leur sont propres, et qui permettraient aux objets de communiquer entre eux. Certaines définitions mettent le point sur les aspects techniques d'IoT tandis que d'autres se concentrent plutôt sur les usages et les fonctionnalités. Néanmoins, nous pouvons considérer la définition proposée par Pierre-Jean Benghozi, Sylvain Bureau et Françoise Massit-Folléa dans leur œuvre "L'internet des objets" et qui se présente comme suit :

« L'Internet des Objets est un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant. ». [3]

1.2. Usages et Fonctionnalités

Les domaines d'utilisation de l'internet des objets semblent être infinis et donnent naissance à des innovations en tout genre. Désormais, les start-ups, les marques et les grandes entreprises innovent dans tous les secteurs pour proposer des nouveaux objets répondant aux besoins de tout le monde. Conscients du potentiel que représente le marché des objets connectés, ils veulent être les premiers à lancer leurs innovations. La figure 4 montre la multitude des secteurs

d'activités qui peuvent bénéficier de la connexion des objets à Internet. En plus, nous citons dans ce qui suit quelques exemples pour montrer l'importance de la connexion intra-machines.

Tableau 1 : Domaines d'application de l'IoT

Secteur d'activité	Domaine d'application
Industrie	Systèmes d'alerte sur les lignes de production
	Contrôle consommation énergétique
	Entrepôts automatisés
	Processus de réapprovisionnement
Commerce	Renvoyer aux acheteurs des promotions personnalisées
	Étiquettes intelligentes pour recevoir des informations sur Smartphones et tablettes,
	Processus de logistique
Services	Paiement sans contact
	Transport : accès au réseau, information
	Santé : télé cardiologie, surveillance de dialyse, surveillance maladies chroniques, remontés d'alarmes, intervention d'urgence
Secteur public	Contrôle et maintenance des équipements : niveau d'eau, présence de bateaux, canalisations d'eau
	Automatisation : parking, bornes de stationnement, systèmes d'arrosage, capteurs pour chauffage
	Systèmes d'alerte opérationnels : alerte de crue, détection de départ de feu de forêt



Figure 4 : Secteurs d'activités de l'Internet des objets

1.3. Les composants de l'IoT

L'Internet des objets n'est pas une technologie mais un système de systèmes. Puisqu'il n'y a pas une décomposition conventionnelle en sous-systèmes, nous avons simplifié le système en six couches que nous estimons importantes. La décomposition en couche nous aide à comprendre l'IoT et mieux positionner notre projet dans ce vaste domaine.

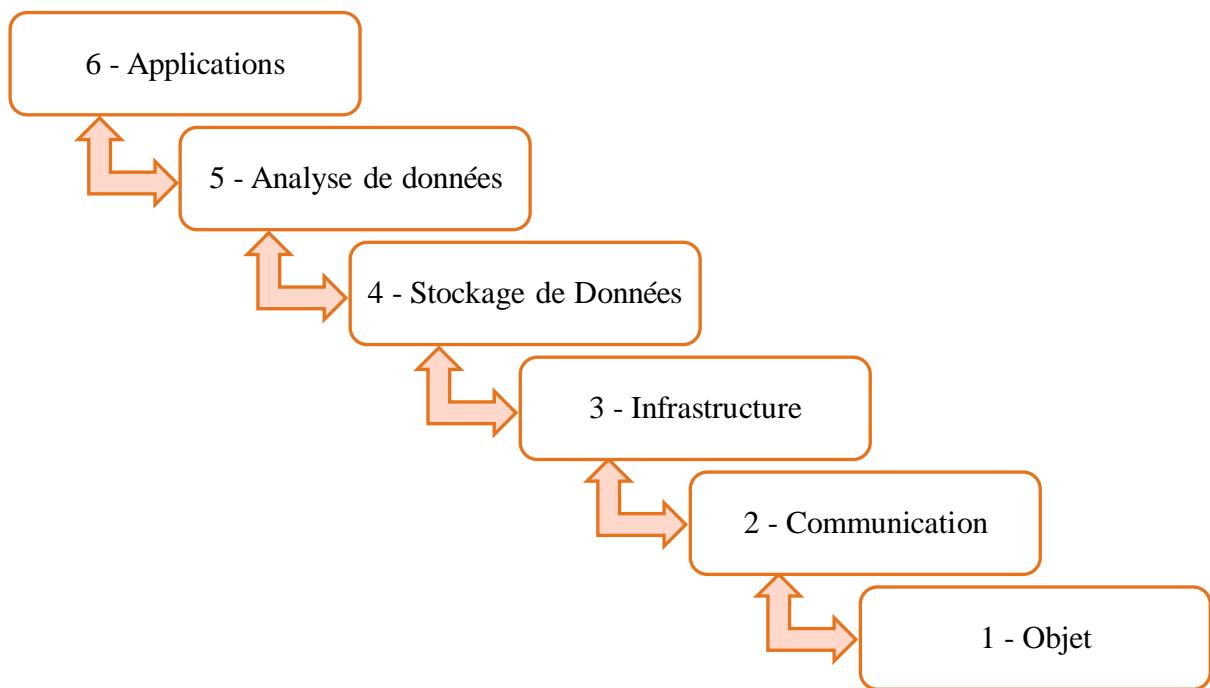


Figure 5 : Les couches logique de l'infrastructure IoT

Notre solution de surveillance réseau s'étale sur les deux dernières couches à savoir l'analyse de données et la couche application clients à travers la plateforme de visualisation des données et les tableaux de bord.

2. Solution Sagemcom pour l’Internet des Objets

Sagemcom a fait le choix de LoRa pour baser son offre de bout en bout pour l’Internet des Objets.

2.1. Définition LoRa

Le réseau LoRa de son nom Long Range ou « longue portée ». Il s’agit d’une technologie qui permet aux objets connectés d’échanger des données de faible taille en bas débit. Comme un premier impact, cela permettra de réduire la consommation énergétique des appareils, leur conférant jusqu’à 10 ans d’autonomie. [4]

Cette technologie utilise à la fois les fréquences radio libre 868 MHz et Internet. Peu gourmande en débit et en énergie, elle a l’avantage d’être très économique pour l’utilisateur final avec une excellente capacité de pénétration des bâtiments, caves et sous-sols. Elle intègre le domaine de l’Internet des Objets par le champ de communication de « Machine to Machine ». [4]

En fait, LoRa, c’est le nom donné à la technologie de modulation du protocole LoRaWAN. Elle a vu le jour en 2012 suite à l’achat par l’entreprise américaine Semtech de la start-up de Grenoble Cycléo. [4]

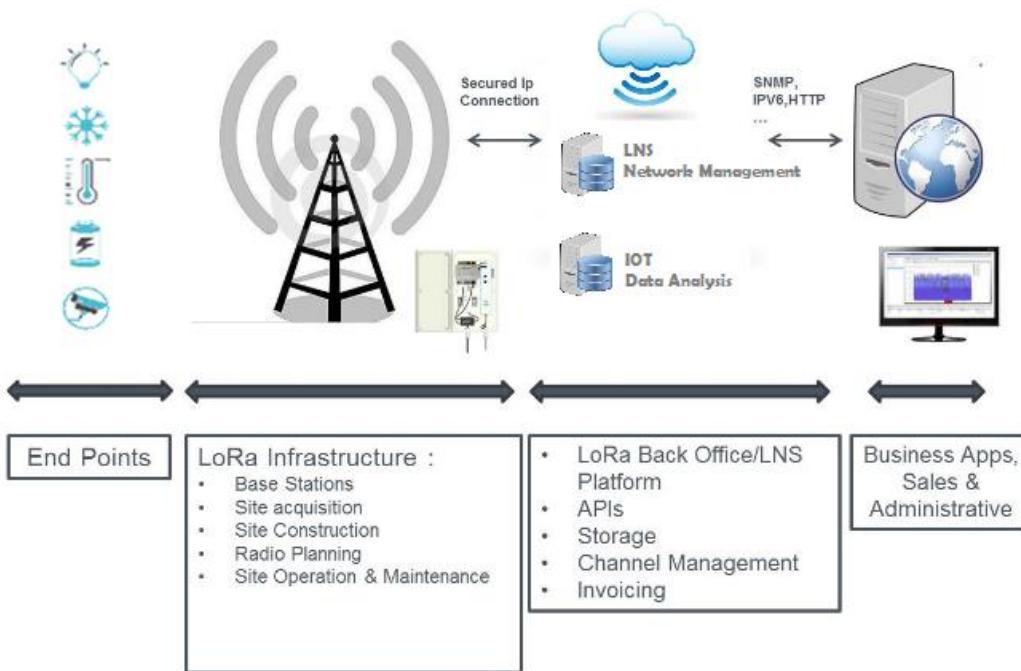


Figure 6 : Architecture réseau de la solution IoT de Sagemcom

2.2. L'alliance LoRa

Annoncée en janvier puis créée en mars 2015, l'Alliance Lora est une association ouverte de leaders d'industriels et d'opérateurs qui ont pour mission de favoriser l'interopérabilité des différents réseaux télécoms et de standardiser les radio-technologies à longue distance appelées LPWAN (Low Power Wide Area Networks).^[4]



Figure 7 : Membres de l'alliance LoRa

Sagemcom, membre fondateur de l'Alliance LoRa, participe ainsi au déploiement de LoRa dans le monde. L'Alliance comprend aujourd'hui plus que 320 membres et couvre les 5 continents.^[5]



Figure 8 : Couverture géographique du réseau LoRa

2.3. Le choix de la technologie LoRa

La technologie LoRa a été retenue par Sagemcom pour tous les atouts qu'elle présente par rapport à ses concurrents. Elle est bidirectionnelle, ce qui permet aux capteurs d'émettre de l'information mais également d'en recevoir. Elle présente également des avantages induits moins visibles à l'utilisateur :

- Le réseau peut adapter le facteur d'étalement, donc le débit et la puissance radio, aux conditions locales de chaque end-point. Ainsi, plus la densité d'antennes réseaux est importante, plus la capacité du réseau augmente ! D'autant que cette technologie est peu atteinte par le bruit ambiant. Deux critères de robustesse très importants pour un opérateur.
- L'optimisation du débit et de la puissance permet d'optimiser également la consommation d'énergie des end-points : avec une consommation énergétique régulée, les objets bénéficient d'une autonomie plus grande !

Elle est interopérable, permettant des modes de communication adaptés à différents types d'objets :

- Tous les objets qui demandent peu d'échanges d'informations mais qui doivent absolument être très autonomes en énergie.
- Tous les objets qui demandent d'être autonomes énergétiquement et qui doivent pouvoir échanger des informations de manière fiable.
- Tous les objets qui doivent être interactifs, et qui doivent pour cela disposer d'une source d'énergie adaptée ou régénérée.

Enfin, la technologie LoRa permettra, moyennant des développements complémentaires, d'offrir un service de localisation d'objets beaucoup moins énergivore que ne le serait l'usage d'un récepteur GPS. ^[6]

3. Missions du stage

Sagemcom, membre fondateur de l'alliance LoRa à forte valeur ajoutée, vise à mettre en place une solution de supervision de la QoS pour le réseau d'Internet des objets LoRa à travers l'analyse des données échangées entre les objets connectés et le serveur LoRa « LNS ».

3.1. Importance de la QoS dans un réseau IoT.

La QoS correspond à la performance globale d'un réseau perçue par les utilisateurs du réseau. Ce concept est très important pour la communication M2M vu que cet échange se fait sans intervention humaine. Il est donc important d'offrir aux utilisateurs du réseau dédié à l'internet des objets LoRa un service de qualité qui couvre mieux leurs objets et qui garantisse un trafic fluide sans perturbation, coupure ou latence.

La qualité de service se mesure et des plans d'actions se planifient pour remédier aux défauts. Notre rôle est de pouvoir informer les superviseurs de ce réseau par les indicateurs de performance réseau.

3.2. Définition des objectifs

L'étude fonctionnelle d'une solution est fondée sur l'ensemble des fonctionnalités et performances que cette dernière offre à ses clients. De ce fait, le système à développer doit garantir un ensemble de besoins que nous allons citer dans la partie qui suit.

La solution de mesure de la QoS doit permettre de dégager les indicateurs de performance clés du réseau. Elle doit aussi être flexible pour tout ajout ou suppression d'indicateurs.

Les résultats de ces analyses doivent être visualisés dans des tableaux de bord. L'application de visualisation doit être multi-utilisateurs avec une bonne gestion des rôles et profils. Elle doit montrer les indicateurs de performance dans des vues globales ainsi que détaillées.

Notre solution doit assurer aussi les objectifs fonctionnels suivants :

- Visualiser les résultats d'analyse en temps réel pour pouvoir prendre les bonnes décisions au bon moment.
- Être compatible avec différentes sources de données avec différents types d'accès et différents formats de messages.
- Mettre en place un système ouvert à d'autres applications autours des données LoRa.
- Supporter un volume massif de données.
- Supporter la monte en charge et la haute disponibilité.

3.3. Les défis liés aux données.

L'architecture de la communication des objets de type LoRa est une diffusion en étoile ce qui signifie que chaque message envoyé par l'objet est capté par toutes les passerelles qui interceptent ce message. Puis chaque passerelle envoie le flux reçu au serveur de réseau LoRa tout en ajoutant ses détails au message.

Le réseau de simulation actuel comporte à peu près 460 objets et 570 passerelles. Nous recueillons en moyenne 150 messages par minute. Les messages reçus sont sous format « JSON » de taille 1.2Ko. Nous collectons donc chaque jour environs 250Mo de données.

Toutes ces données semblent être gérables par les systèmes de gestion des données habituelles. Mais, ce n'est qu'un environnement de test. Sagemcom vise avoir 30 millions d'objets d'ici jusqu'à l'année 2020. Alors que si chaque objet envoie un seul message par jour et que ce message est intercepté par trois passerelles seulement, nous serons face à 144 Go de données par jour. Aucun système de gestion de base de données relationnelle n'est capable de gérer ce volume énorme de données.

Nous aurons forcément recours aux nouvelles technologies du Big Data qui eux seules sont capables d'intercepter ce flux de données. Nous devons prendre en considération le volume des données massif dans notre architecture du système.

Conclusion

Dans ce deuxième chapitre, nous avons eu un petit aperçu sur l'Internet des objets et sur l'offre de Sagemcom dans ce domaine. Nous avons aussi détaillé les différents objectifs de notre solution.

Nous présenterons dans le chapitre suivant l'architecture de la solution proposée.

CHAPITRE 3 :

ARCHITECTURE DE LA

SOLUTION PROPOSÉE

Introduction

La première contrainte à prendre en considération lors de la conception de notre système est la taille massive des données générées par les objets connectés. Nous devrons donc utiliser des technologies spécifiques capables de manipuler des grandes quantités d'information, des architectures « Big Data ».

L'horizon de l'Internet des objets et du « Big data » ouvre la possibilité de connecter les personnes ou les objets d'une manière plus pertinente, de fournir la bonne information au bon destinataire et au bon moment, ou encore de faire ressortir les informations utiles à la prise de décision.

1. Introduction au « Big Data »

1.1. Définition

Le « Big Data », littéralement les Grosses données, est une expression anglophone utilisée pour désigner des ensembles des données qui deviennent tellement volumineux que nous ne pouvons pas les manipuler avec des outils classiques de gestion de base de données.^[7]

Dans ces nouveaux ordres de grandeur, la capture, le stockage, la recherche, le partage, l'analyse et la visualisation des données doivent être redéfinis.^[7]

Il s'agit donc d'un ensemble de technologies, d'architecture, d'outils et de procédures permettant à une organisation de capter rapidement, traiter et analyser de larges quantités et contenus hétérogènes et changeants, et d'en extraire les informations pertinentes à un coût accessible. Le « Big Data » couvre 3 dimensions importantes : **Volume, Variété, Vitesse** :

- Le volume : Correspond à la masse d'informations produite chaque seconde. Selon des études, pour avoir une idée de l'accroissement exponentiel de la masse de données, nous considérons que 90 % des données ont été engendrées durant les années où l'usage d'internet et des réseaux sociaux a connu une forte croissance. L'ensemble de toutes les données produites depuis le début des temps jusqu'à la fin de l'année 2008, conviendrait maintenant à la masse de celles qui sont générées chaque minute. Dans le monde des affaires, le volume de données collectées chaque jour est d'une importance vitale.
- La variété : Seulement 20% des données sont structurées puis stockées dans des tables de bases de données relationnelles similaires à celles utilisées en gestion comptabilisée.

Les 80% qui restent sont non-structurées. Cela peut être des images, des vidéos, des textes, des voix, et bien d'autres encore... La technologie « Big Data », permet de faire l'analyse, la comparaison, la reconnaissance, le classement des données de différents types comme des conversations ou messages sur les réseaux sociaux, des photos sur différents sites etc. Ce sont les différents éléments qui constituent la variété offerte par le « Big Data ».

- La vitesse : c'est la rapidité de l'élaboration et du déploiement des nouvelles données. Il s'agit d'analyser les données au décours de leur lignée (appelé parfois analyse en mémoire) sans qu'il soit indispensable que ces informations soient entreposées dans une base de données. [8]

1.2. Les technologies « Big Data »

Les créations technologiques qui ont facilité la venue et la croissance du « Big Data » peuvent globalement être catégorisées en deux familles : d'une part, les technologies de stockage. D'autre part, l'arrivée de technologies de traitement ajustées.

La première solution permet d'implémenter les systèmes de stockage considérés comme plus performants que le SQL traditionnel pour l'analyse de données en masse. Ce sont les bases de données NoSQL avec leurs différents types (orienté clé/valeur, document, colonne ou graphe).

La deuxième solution est appelée le traitement massivement parallèle. Le Framework Hadoop en est un exemple. Celui-ci combine le système de fichiers distribué HDFS et l'algorithme MapReduce.

D'autres applications qui enrichissent l'écosystème ont vu le jour. Chacune d'entre elles peut être classifiée parmi une ou plusieurs catégories de cette liste :

- Ingestion/Extraction de données,
- Traitement de données,
- Analyse/Apprentissage,
- Data visualisation,
- Requête/Interrogation,
- Workflow,
- Stockage,

- Ordonnancement,
- Sécurité,
- Gouvernance,
- Messages. [9]

L’interaction entre les différentes catégories de cet écosystème est représentée dans l’architecture globale ci-dessous.

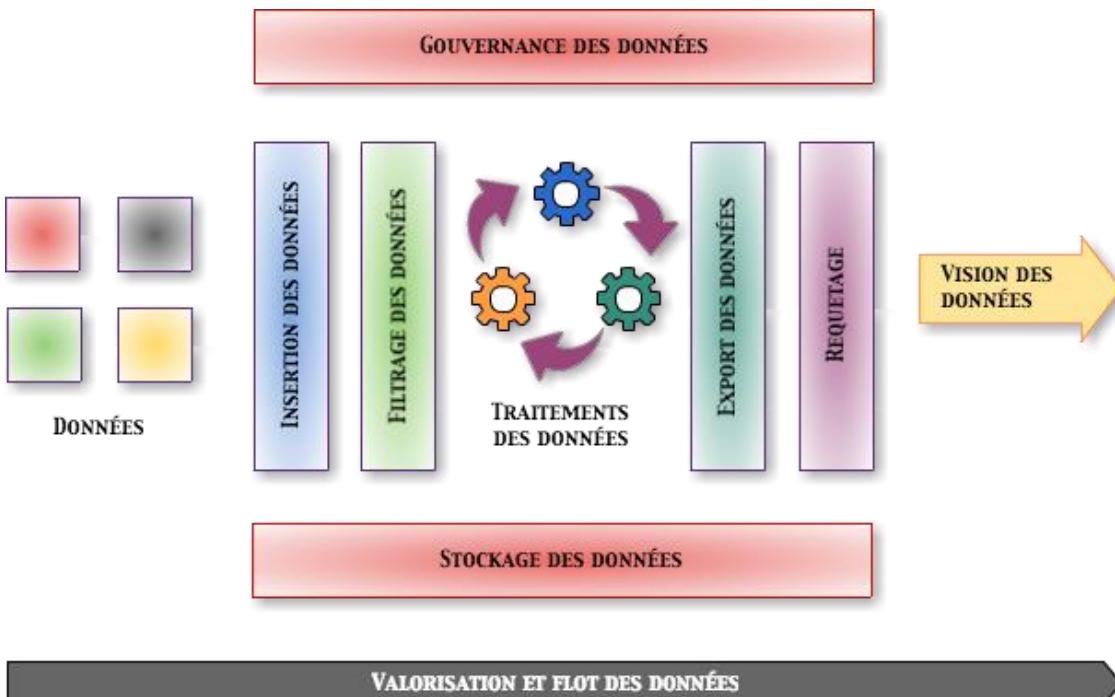


Figure 9 : Architecture d'une solution Big Data

1.3. Définition de Hadoop

Hadoop est un Framework open source, écrit en Java et géré par la fondation Apache. Il est capable de stocker et traiter de manière efficace un grand nombre de données, en reliant plusieurs serveurs banalisés, appelé cluster, entre eux pour travailler en parallèle. Il gère la distribution des données au cœur des machines du cluster, leurs éventuelles défaillances, mais aussi l’agrégation du traitement final. L’architecture est de type « Share nothing » : aucune donnée n’est traitée par deux nœuds différents, même si les données sont réparties sur plusieurs nœuds. [9]

Hadoop est composé de quatre éléments :

- **Hadoop Distributed File System (HDFS)** : un système de fichiers distribué pour le stockage persistant des données,

- **Hadoop YARN** : un Framework de gestion des ressources et de planification des traitements,
- **Hadoop Common** : ensemble d'utilitaires utilisés par les autres briques Hadoop,
- **Hadoop MapReduce v2** : Un Framework de traitements distribués basé sur YARN.

Les points forts d'Hadoop sont :

- Distribution des traitements au plus près de la donnée.
- Reprise automatique sur erreur.
- Coût de stockage très concurrentiel.
- Écosystème très important.

Ses inconvénients sont :

- Performance (par rapport à des bases NoSQL, à des grilles de données, etc.).
- Fait principalement pour analyser de très gros volume de données.

1.4. Définition des bases de données NoSQL

NoSQL n'est pas à prendre forcément comme « not » SQL, mais plutôt « not only ». Le NoSQL désigne une famille de systèmes de gestion de base de données qui s'écarte du paradigme classique des bases relationnelles. Il propose des solutions qui peuvent venir compléter les solutions des SGBDR classiques.

Le NoSQL englobe une gamme étendue de technologies et d'architectures, afin de résoudre les problèmes de performances en matière d'évolutivité et de Big Data que les bases de données relationnelles ne sont pas conçues pour affronter.

Le NoSQL est particulièrement utile lorsqu'une entreprise doit accéder, à des fins d'analyse, à de grandes quantités de données non structurées ou de données stockées à distance sur plusieurs serveurs virtuels du Cloud. ^[10]

Le NoSQL propose plusieurs formes de stockage de l'information :

- **Clé/valeur** : à la manière des tableaux associatifs des langages de programmation.
A une clé correspond une valeur.
- **Orienté colonne** : à une clé correspond un ensemble de colonne, chacune ayant une valeur.

- **Orienté document** : à une clé correspond des ensembles champs/valeurs qui peuvent être hiérarchisés
- **Orienté graphe** : les données sont modélisées sous forme de nœuds qui ont des liaisons entre eux.

2. Étude et choix techniques

2.1. Objectifs de l'étude

Comme nous avons vu dans la première partie de ce chapitre, le nombre de solutions qui peuvent répondre à nos besoins est un nombre considérable à première vue. De plus, l'écosystème est très riche par des modules de différentes catégories. Dans le cadre de l'anticipation, nous devons étudier les différentes solutions disponibles pour trouver la meilleure architecture qui répond au mieux à nos besoins.

L'objectif de notre stage consiste à déployer une solution pour superviser la QoS du réseau LoRa. Il faut mettre en place une architecture à la fois ouverte et générique autour des données récupérées depuis les objets. Cela servira pour mesurer les indicateurs de performance réseau et les futures applications à mettre en place pour répondre aux nouveaux besoins.

2.2. Étude comparative Hadoop et NoSQL

Nous avons vu que les outils Big Data sont reparties entre deux technologies la plateforme Hadoop et les bases de données NoSQL. Nous allons donc étudier les points de ressemblance et de distinction entre ces deux technologies pour décider laquelle utiliser.

2.2.1. Les points de ressemblance

Nous notons plusieurs points de ressemblance entre le Framework Hadoop et les bases de données NoSQL dans le fait que les deux solutions peuvent gérer des volumes de données importants et en rapide augmentation. Ils sont particulièrement efficaces avec des formats de données hétérogènes, même si ces données évoluent au fil du temps.

Pour plus de précisions sur la gestion du volume de données, les deux technologies peuvent utiliser du matériel informatique de base travaillant ensemble comme un cluster. Pour traiter de plus grands ensembles de données, il suffit d'ajouter plus de matériel au cluster dans un modèle connu sous l'échelle horizontale, également désigné comme « Scaling Out ». Comparez cela au

modèle connu sous l'échelle verticale, dans lequel nous mettons à jour nos serveurs existants avec du matériel informatique plus puissant.

En ce qui concerne les formats de données, les deux technologies sont appropriées pour les différents types de données générées par nos objets connectés.

2.2.2. Paramètres distinctifs

Vu la forte ressemblance entre ces deux technologies, il semble que NoSQL et Hadoop soient en concurrence directe. Mais, bien qu'ils soient conçus pour le Big Data, ces deux environnements se distinguent par les types de charges de travail qui leur sont destinées. En effet, la technologie NoSQL assure généralement un accès interactif et en temps réel aux données avec des scénarios d'utilisation qui impliquent des interactions avec l'utilisateur, telles que des applications Web. L'avantage principal, est la capacité de lecture et d'écriture des données très rapide.

Pour sa part, Hadoop traite d'énormes quantités de données. Pour ce faire, les données sont réparties en différents nœuds du cluster, tandis que le traitement est généralement effectué simultanément par plusieurs serveurs. Hadoop procède à cette répartition du travail en utilisant la méthodologie MapReduce. Dans la mesure où chaque serveur comprend une partie de l'ensemble du jeu de données, MapReduce rapproche le traitement au plus près possible des données afin d'éviter de ralentir l'accès au réseau par des opérations de transfert de données.

2.2.3. Le choix de la technologie

Finalement, NoSQL et Hadoop travaillent mieux ensemble comme composants dans une architecture de données d'entreprise unifiée. En fait, ces deux technologies sont complémentaires et nous pouvons les déployer ensemble dans une architecture unifiée. Nous pouvons utiliser NoSQL pour gérer les données des capteurs et ensuite utiliser Hadoop pour analyser ces données pour générer des recommandations, effectuer des analyses prédictives, ou détecter des activités frauduleuses.

Tableau 2 : Tableau comparatif des technologies Big Data

	NoSQL	Hadoop
Traitement	Temps réel	Batch
Usage	Applications Interactives Lecture/écriture rapides	Traitement grande échelle Grande puissance de calcul
Point Fort	Amélioration des performances à l'échelle horizontale	
Type de donnée	Différents type des données	

2.3. Architecture Globale

Nous savons bien maintenant que nous avons besoins des deux technologies pour réussir notre solution et aboutir à des résultats qui répondent au mieux aux besoins définis dès le début. Nous sommes aussi intéressés par les bases de données NoSQL multidimensionnelles qui intègrent le traitement analytique en ligne « OnLine Analytical Processing », « OLAP » en anglais, pour l'analyse et le calcul des indicateurs de performance réseau. Ces modules ressemblent aux moteurs OLAP connus dans le monde relationnel sauf qu'ils sont adaptés au volume massif des données.

L'architecture globale de notre système est décrite dans la figure ci-dessous :

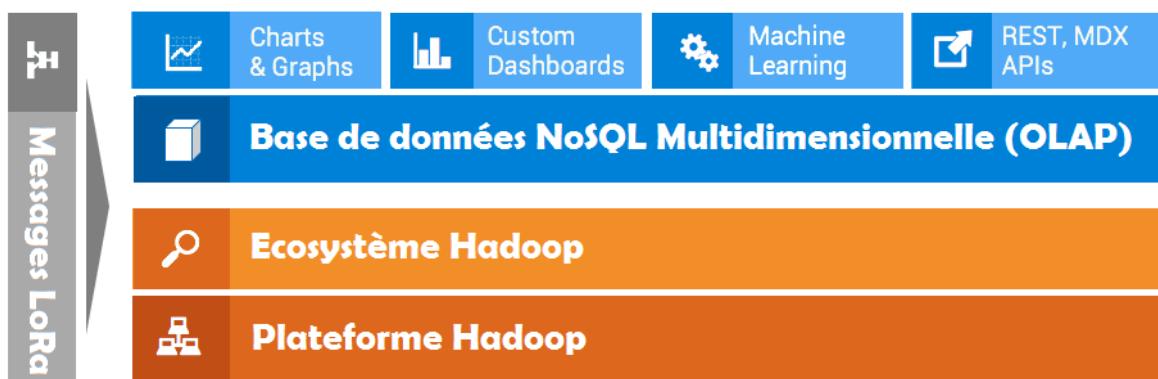


Figure 10 : Architecture globale du système

2.4. Choix de la distribution Hadoop

2.4.1. Les avantages de l'utilisation d'une distribution Hadoop.

Hadoop est un projet apache open source, son installation en local sur un système est assez simple. Mais, elle est plus complexe sur plusieurs nœuds (configuration avancée, gestion de droits...).

L'écosystème Hadoop est très riche en modules et services ce qui rend la tâche d'administration plus difficile. Il est, donc, recommandé d'utiliser une distribution Hadoop.

Une distribution contient différents projets de l'écosystème Hadoop. Ceci assure que toutes les versions utilisées fonctionnent ensemble sans problèmes. Ces packages sont offerts par des fournisseurs de la Plateforme.

Aujourd'hui nous distinguons trois distributions Hadoop leaders dans le marché des solutions Big Data à savoir Hortonworks, Cloudera et MapR. En plus du package, ces fournisseurs de distribution offrent des outils graphiques pour le déploiement, l'administration et le monitoring des clusters Hadoop. De cette façon, il est beaucoup plus facile d'installer, de gérer et de surveiller les clusters.

2.4.2. Hortonworks Data Platform

Hortonworks Data Platform est la seule plateforme de données Apache Hadoop 100 % open source et elle est la distribution Hadoop la plus proche du projet Apache initial. L'équipe de Hortonworks, qui travaille sur cette plateforme, représente la majorité de la communauté de l'écosystème Hadoop : au sein de cette communauté, ils représentent et pilotent les exigences de l'entreprise au sens large : YARN, élément essentiel de la plateforme Hadoop 2.0, est initialement un développement Hortonworks. [11]

Les entreprises peuvent se servir de la plateforme de données Hortonworks pour bâtir et déployer des applications basées sur Hadoop. Hortonworks a des relations approfondies avec les principaux fournisseurs de centres de données stratégiques, ce qui permet aux utilisateurs de bénéficier de ces produits dans le cloud.

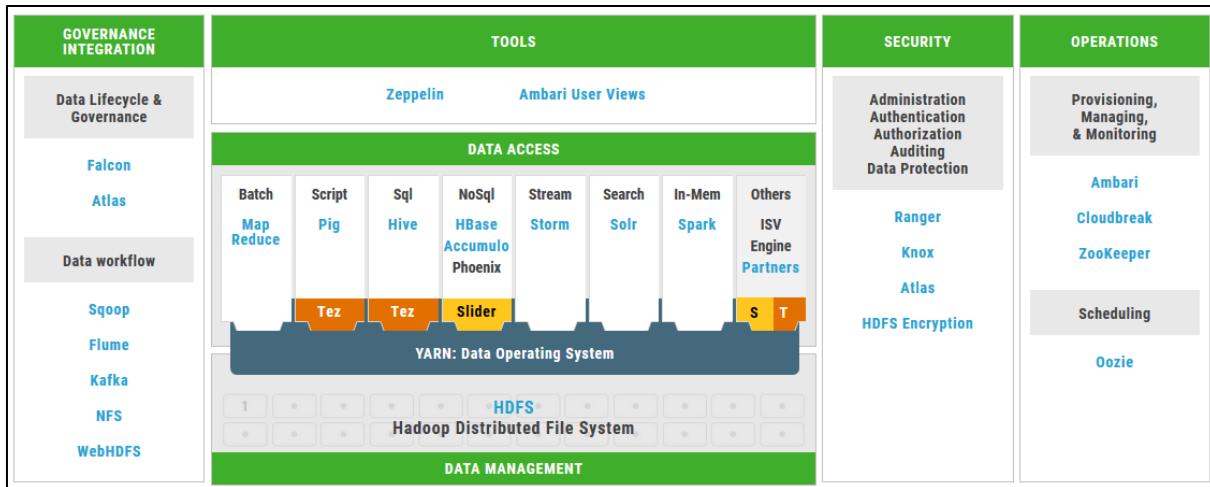


Figure 11 : Architecture de la distribution Hortonworks

2.4.3. Cloudera

Cloudera a été créé en 2008. Ce fut la première entreprise à développer et à distribuer Hadoop, et c'est aujourd'hui la plateforme qui compte le plus grand nombre d'utilisateurs. Elle est facile à déployer grâce à son gestionnaire graphique Cloudera Manager, cette distribution se distingue par un travail très important effectué sur les outils d'administration. [11]

Cloudera est également à l'origine des outils suivants :

- **Sqoop**, qui permet d'importer des données depuis des bases de données relationnelles dans le système de fichier HDFS.
- **Flume**, qui permet de collecter et intégrer des grands volumes de fichiers de log dans Hadoop.
- **Impala**, qui permet de requêter les données contenues dans le cluster via SQL.

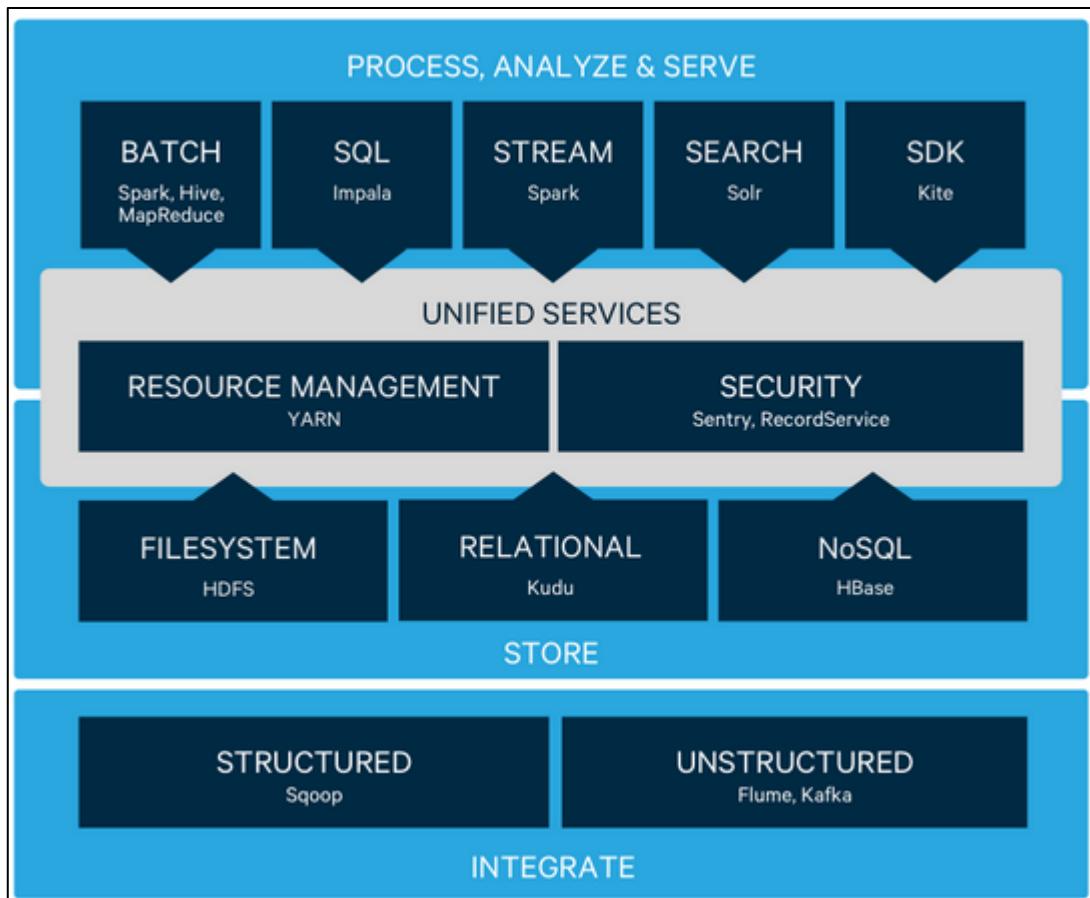


Figure 12 : Architecture de la distribution Cloudera

2.4.4. MapR

Bien que son approche soit commerciale, MapR contribue à des projets Apache Hadoop comme HBase, Pig, Hive, ZooKeeper et Drill. MapR se distingue surtout de la version d'Apache Hadoop par sa prise de distance avec le cœur de la plate-forme. MapR propose ainsi son propre système de fichiers distribué ainsi que sa propre version de MapReduce : MapR FS et MapR MR. La solution propriétaire de MapR reste fidèle à l'écosystème Hadoop comme le montre l'architecture de la plateforme.^[11]

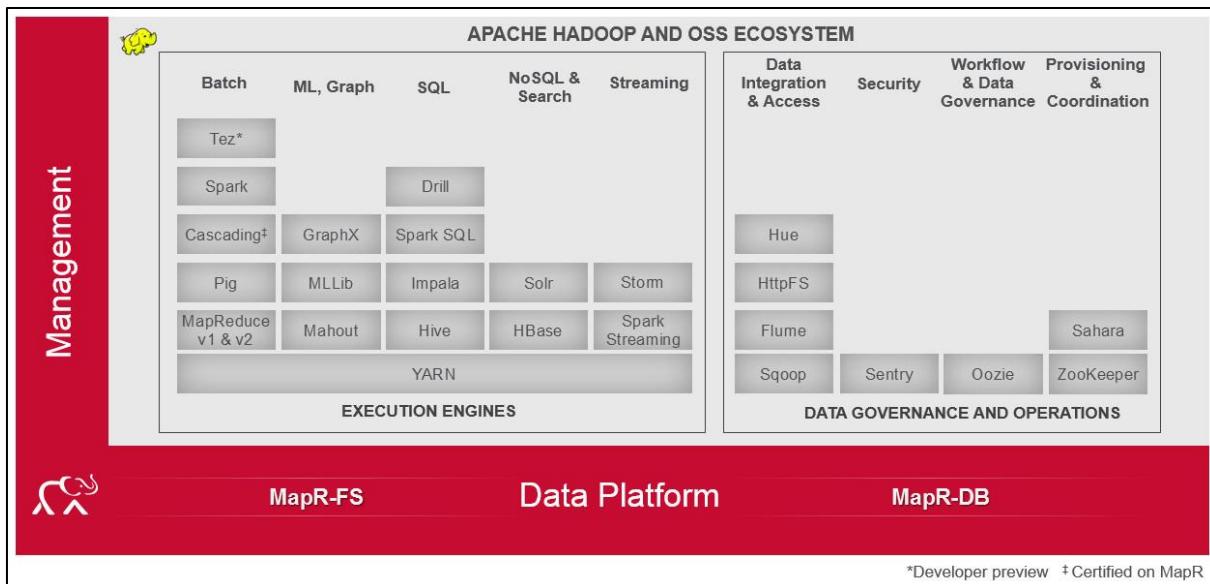


Figure 13 : Architecture de la distribution Map-R

2.4.5. Étude Comparative des distributions Hadoop

Nous allons nous appuyer pour la comparaison des distributions sur l'étude faite par « Robert D. Schneider » dans son guide des acheteurs des distributions Hadoop faite en 2004 dont le résumé est décrit dans le tableau suivant :

Tableau 3 : Tableau comparatif des distributions Hadoop

	Hortonworks	Cloudera	MapR
Performance et Monte en charge			
Intégration des données	Batch	Batch	Batch et lecture du flux
Architecture des métadonnées	Centralisée	Centralisée	Distribuée
Performance HBase	Pics de latence	Pics de latence	Faible latence
Applications NoSQL	Applications batch	Applications batch	Applications batch et online/real-time
Fiabilité			
Haute disponibilité	Récupération d'une erreur à la fois	Récupération d'une erreur à la fois	Récupération multiple d'erreurs

MapReduce	Redémarrage des jobs	Redémarrage des jobs	Sans redémarrage
Mise à jour	Arrêt planifié	Mise à jour à chaud	Mise à jour à chaud
RéPLICATION	Données	Données	Données et métadonnées
Sauvegarde	Seulement pour les fichiers fermés	Seulement pour les fichiers fermés	Pour tous les fichiers
Reprise après sinistre	Non	Copie planifiée des fichiers	Miroir
Gestion			
Outils de gestion du cluster	Ambari	Cloudera Manager	MapR Control System
Volume Support	Non	Non	Oui
Alarmes Alertes	Oui	Oui	Oui
Intégration avec REST API	Oui	Oui	Oui
Contrôle des données et des jobs	Non	Non	Oui
Accès aux Données			
Accès aux fichiers système	HDFS, NFS lecture seule	HDFS, NFS lecture seule	HDFS lecture/écriture, NFS
I/O Fichier	Ajout seulement	Ajout seulement	Lecture/écriture
Sécurité : ACLs	Oui	Oui	Oui
Authentification Réseau	Kerberos	Kerberos	Kerberos, Native

Robert D. Schneider, Hadoop Buyer's Guide, Ubuntu, 2014

2.4.6. Le choix de la distribution Hadoop

Les trois distributions ont une approche et un positionnement différent en ce qui concerne la vision d'une plateforme Hadoop. Nous retenons alors une solution open source pour notre

solution. Hortonworks répond le mieux à nos besoins actuels et nous allons l'utiliser pour la suite de nos travaux comme distribution Hadoop.

2.5. Choix de la base de données NoSQL

2.5.1. Les bases de données cibles

Nous avons fait le choix de focaliser notre étude sur les modules d'archivage de données qui permettent de traiter les données LoRa dans des cubes OLAP. La technologie OLAP est derrière la plus part des solutions d'aide à la décision existantes dans le monde relationnel. Ces cubes sont utilisés principalement pour afficher et analyser les données sur les multiples dimensions. Dans le monde du Big Data, ces systèmes sont adaptés à gérer la quantité massive de données d'une manière performante et évolutive ce qui permet aux utilisateurs d'interroger et d'analyser de grands volumes d'information en temps réel.

Dans l'écosystème Hadoop il existe plusieurs modules qui répondent à nos exigences. Mais, nous sommes intéressés par les offres open source du marché dont deux solutions en particulier : Apache Kylin et Druid. Nous allons étudier chacune à part pour pouvoir les comparer et choisir l'une d'entre elles pour l'utiliser dans notre système.

2.5.2. Druid

Druid est un entrepôt de données open source conçu pour les requêtes OLAP sur des données d'événementielles. Druid fournit une faible latence d'intégration, d'exploration et d'agrégation de données. Des déploiements existants de Druid ont pu supporter des milliards d'événements et des pétaoctets de données.^[12]

L'ensemble de données est composé de trois éléments distincts :

- Colonne Temps : Cette colonne est traitée séparément, car l'ensemble des recherches requis sont autour de l'axe du temps.
- Colonnes Dimension : Les dimensions sont les attributs de chaîne d'un événement, et les colonnes les plus utilisées dans le filtrage des données.
- Colonnes métriques : Les métriques sont des colonnes utilisées dans les agrégations et les calculs. Les métriques sont généralement des valeurs numériques et des calculs incluent des opérations telles que le comptage, la somme, et la moyenne.

Parmi les principales caractéristiques de Druid :

- **Des réponses en mois d'une seconde** pour des requêtes OLAP avec le filtrage et l'analyse exacte de ce qui est nécessaire pour une requête multidimensionnelle complexe. Agrégation et filtrage sur les données en millisecondes.
- **Intégration des données en temps réel et en continu.** Druid emploie l'ingestion sans verrou de données pour permettre l'ingestion et l'interrogation de plus de 10.000 événements par seconde par nœud simultanée. Autrement dit, le temps de latence entre le moment où un événement se produit et quand il est visible est limité par la rapidité avec laquelle l'événement peut être livré à Druid.
- **Applications Analytiques très puissantes.** Druid a de nombreuses fonctionnalités intégrées pour pouvoir servir des applications analytiques conçues pour être utilisées par des milliers d'utilisateurs simultanés.
- **Druid est hautement disponible.** Il prend en charge les mises à jour sans arrêt afin que les données soient toujours disponibles et interrogables au cours des mises à jour logicielles.
- **Druid est hautement évolutif.** Des déploiements existants gèrent des milliards d'événements, pétaoctets de données, et des milliers de requêtes par seconde.

2.5.3. Apache Kylin

Kylin est une solution initialement développée par eBay pour ses besoins internes et qui est aujourd’hui ouverte à tous. Le fonctionnement interne de Kylin inclut des méthodes pour stocker les résultats pré-calculés pour servir les requêtes d'analyse, de générer des cubes de données de chaque niveau avec toutes les combinaisons possibles de dimensions et de calculer tous les indicateurs à différents niveaux. [13]

La plateforme Kylin offre donc des caractéristiques importantes pour l'analyse Big Data :

- Un moteur OLAP très rapide conçu pour réduire le temps de latence d'Hadoop pour les données de plus de 10 milliards de ligne.
- Le support ANSI SQL sur l'interface Hadoop.
- Capacité de requêtes interactives sur Hadoop via Kylin en moins d'une seconde
- Des cubes de requête MOLAP avec une pré-construction dans Kylin au-delà de 10 milliards d'enregistrements bruts.
- L'intégration transparente avec les outils de BI comme Tableau.

- Un driver ODBC open source entièrement développé pour Kylin interagissant avec Tableau.

2.5.4. Le choix de l'entrepôt de données.

Apache Kylin est une solution plus complète que Druid. Il offre une panoplie de connecteurs pour le monde SQL et les outils de visualisations existantes. Il a l'inconvénient de recalcul du cube de données, ce qui augmente le temps de disponibilité de l'information et augmente la taille de données par la sauvegarde des résultats d'agrégations.

Druid, par contre, il est moins mature que Kylin mais il est bien adapté pour notre type de données évènementielles. Il fonctionne aussi en mémoire ce qui lui donne une meilleure position pour l'analyse en temps réel. De plus, Druid permet de réduire considérablement la taille de données brutes reçues par les objets.

Nous allons retenir Druid comme entrepôt de donnée et moteur de requêtes OLAP pour notre solution de calcul des indicateurs de performance réseau.

3. Entrepôt de donnée Druid

3.1. Architecture d'un cluster Druid

Un cluster Druid est composé de différents types de nœuds et chaque type de nœud est conçu pour effectuer des tâches bien spécifiques. Les différents types de nœuds fonctionnent assez indépendamment les uns des autres et il y a peu d'interaction entre eux. Par conséquent, les pannes de communication inter-cluster ont un impact minimal sur la disponibilité des données. Les différents types de nœuds sont réunis pour former un système complet pour la résolution des problèmes d'analyse de données complexes. [12]

Les types de nœuds qui existent dans un cluster Druid sont :

- « **Historical** » : Les nœuds Historiques forment généralement le noyau d'un cluster Druid. Les nœuds historiques persistent les segments localement et exécutent les requêtes sur ces segments. Les nœuds ont une architecture distribuée sans échange de données permet de charger des segments, supprimer des segments, et servir des requêtes sur ces segments.

- « **Broker** » : Les nœuds courtiers sont ceux que les clients et les applications les requêtent pour obtenir des données de Druid. Ils sont responsables de la diffusion des requêtes et de la collection et de la fusion des résultats.
- « **Coordinator** » : Les nœuds coordinateurs gèrent des segments sur des nœuds historiques dans un cluster. Les coordinateurs annoncent aux nœuds historiques la création de nouveaux segments, le déposition des anciens segments, et déplacement des segments pour équilibrer la charge.
- « **Overlord** » : Les nœuds d'indexation gèrent les tâches d'indexation des données en mode batch ou temps réel.
- « **Real-Time** » les nœuds temps réel encapsulent la fonctionnalité d'ingérer et servir le flux d'événements en temps réel. Les événements indexés par ces nœuds sont immédiatement disponibles pour requêtes.

Druid a besoin de trois dépendances externes pour les opérer :

- **Zookeeper** : Druid utilise Zookeeper pour la communication intra-cluster.
- **Metadata Storage** : Druid repose sur un système de gestion de base de données relationnelle pour stocker la configuration et les métadonnées sur les segments. Les services qui créent des segments écrivent de nouvelles entrées dans l'entrepôt de métadonnées et les nœuds coordinateurs surveillent cet entrepôt pour savoir quand de nouvelles données doivent être chargées ou anciennes données doivent être abandonnées.
- **Deep storage** : c'est un système de stockage de fichier de préférence distribué pour supporter le volume massif des données. Il sert à sauvegarder d'une manière permanente des segments. Les services qui créent les segments les persistent dans cet entrepôt et les nœuds historiques les retrouvent après.

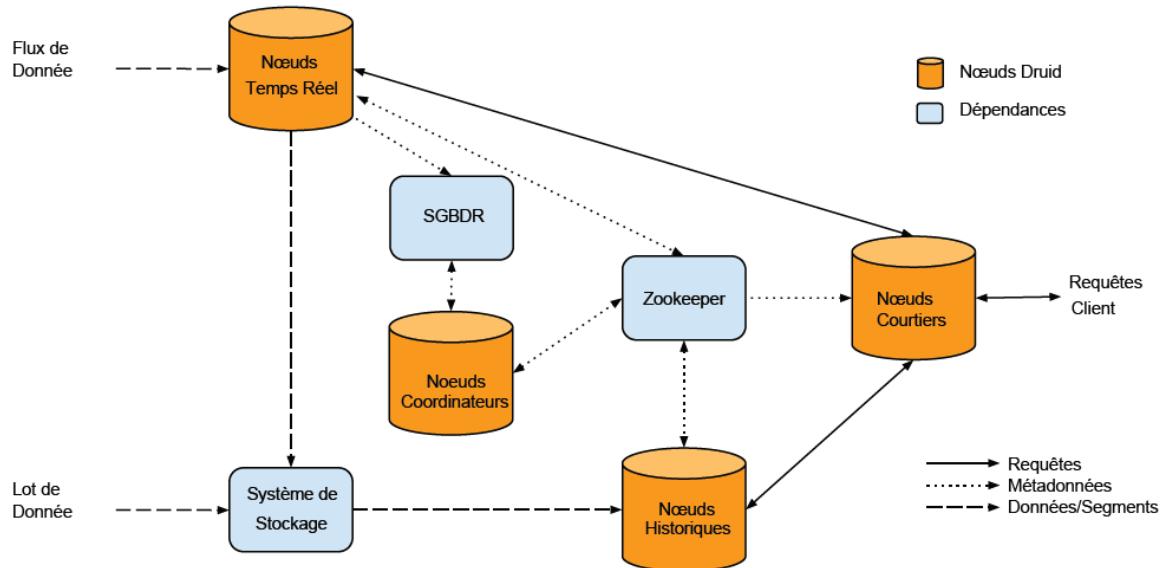


Figure 14 : Architecture globale d'un cluster Druid

3.2. Techniques de stockage

Les sources de données dans Druid sont des collections de données évènementielles historiées et partitionnées en un ensemble de segments. Ces segments sont l'output du processus d'indexation. Ils représentent les structures fondamentales de stockage des données dans Druid. Et ils contiennent les différentes dimensions et mesures dans un ensemble de données sur une période spécifique du temps. Druid stocke les données dans un format orienté colonne personnalisée hautement optimisées pour les agrégations et les filtres.

4. Indexation en temps réel

Notre but est d'aboutir à une solution d'analyse de qualité de service en temps réel comme aborder dans la spécification des objectifs. Par ce choix, nous avons fixé la barre des exigences à son haut niveau. Le but de cela est d'ajouter de la valeur à notre solution et pour anticiper les besoins des clients dans ce sens.

La réussite de cette fonctionnalité est dépendante de la résolution de ces trois obstacles :

- L'acquisition des nouveaux messages.
- Faire les traitements nécessaires pour ce flux
- Charger les données dans Druid.

Tous ces obstacles affectent d'une manière directe le temps d'indexation des données. En augmentant le volume de données en input, la tâche devient alors plus pénible. Dans cette partie, nous allons essayer de construire notre pipeline de données temps réel à l'aide du module d'intégration temps réel de Druid et l'écosystème de Hadoop.

4.1. Livraison des messages avec Kafka

Afin de garantir l'acquisition des messages LoRa avec des faibles latences, nous avons intérêt à utiliser un bus de message de haut débit. Dans l'écosystème Hadoop, le module Apache Kafka est un système très reconnu pour la manipulation de flux de données en temps réel.

Kafka est un système de messagerie rapide, évolutif, durable et tolérant aux pannes qui fonctionne en mode publier/abonner. Kafka est souvent utilisé à la place de courtiers de messages traditionnels comme JMS et AMQP en raison de son plus grand débit, la fiabilité et la réplication. Il fonctionne en combinaison des moteurs de traitement de flux pour l'analyse en temps réel et le rendu des données en continu. [14]

Kafka sera le point de départ de notre pipeline des données temps réel.

4.2. Traitement en temps réel

Le deuxième obstacle rencontré est la manière de traitement des messages. En effet nous devons récupérer les messages depuis le service Kafka afin de les nettoyer, les formater et les envoyer vers le service d'indexation de Druid. Pour ce faire, il nous faut un système de traitement de flux en temps réel. Dans l'écosystème Hadoop nous avons trouvé deux solutions qui peuvent nous intéresser qui sont « Spark Streaming » et « Storm ». Nous allons alors les étudier.

4.2.1. Spark Streaming

Apache Spark est un Framework de traitements Big Data open source construit pour effectuer des analyses. Spark streaming est un module de la plateforme Spark. Il exploite la capacité de Spark par la planification rapide des tâches de base pour effectuer des analyses de flux en temps réel. [15]

4.2.2. Storm

Storm est un Framework de traitement distribué de données en temps réel. Chaque application déployée est sous forme de topologie, chaque nœud envoie le message vers le nœud suivant avec un message traité à la fois. Ces nœuds sont écrits d'une manière personnalisée pour implémenter la logique métier de la topologie.^[16]

4.2.3. Choix du moteur de traitement

Spark Streaming ne fait pas du vrai temps réel mais plutôt du micro-batch c'est-à-dire que Spark ne traite pas les données d'un flux un par un comme Storm, mais elles sont accumulées avant d'être traitées.

Pour notre solution nous allons utiliser Storm car il est mieux adapté pour le traitement en temps réel que Spark streaming qui fonctionne en mode temps quasi réel.

5. Architecture détaillée du système

Notre solution est conçue pour fonctionner avec plusieurs sources de données client. Après la lecture et le nettoyage des messages inutiles, les données sont inscrites dans le module Kafka en attendant leur transfert vers le service de traitement temps réel Storm.

Storm récupère les messages depuis Kafka un par un et il procède à l'intégration de ces messages dans Druid à l'aide du service d'indexation et à la persistance du message sous son format brute dans HBase ou HDFS.

L'application de visualisation fait appel aux résultats d'analyse via des requêtes à l'entrepôt de données Druid.

L'architecture détaillée est présentée dans la figure 15.

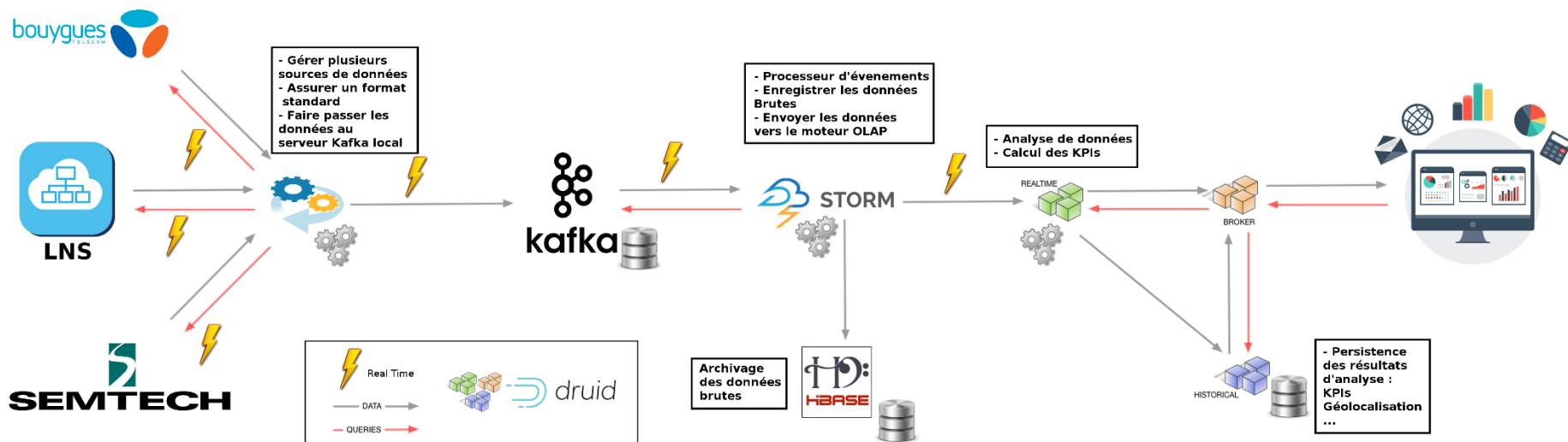


Figure 15 : Architecture détaillé de la solution de supervision de la qualité de service

Conclusion

Nous avons étudié dans ce chapitre les différentes alternatives qui peuvent répondre aux besoins de notre solution pour enfin aboutir à une architecture complète de la solution proposée. Nous allons présenter dans le chapitre suivant les indicateurs de performances qui permettent la supervision de la qualité de service du réseau ainsi que les objectifs de chaque tableau de bord.

CHAPITRE 4 :

IDENTIFICATION DES

INDICATEURS DE

PERFORMANCE

Introduction

Après avoir dégagé les différentes fonctionnalités attendues pour notre solution, nous allons définir les indicateurs de performance responsables de l'évaluation de la QoS du réseau LoRa, ainsi que la présentation des utilisateurs de la solution et les objectifs derrière leurs tableaux de bord.

1. Les indicateurs de performance

Après définition des objectifs, nous commençons maintenant à étudier les indicateurs de performances qui permettent de mesurer la QoS de notre réseau IoT en fonction des besoins et de l'activité exercée. La réussite ou l'échec de notre système d'analyse temps réel dépend des mesures prises au cours de cette analyse. Il faut donc bien réfléchir posément à ce que nous allons mesurer.

Pour le faire il faut suivre une démarche à 3 étapes qui permet de :

1. Identifier les objectifs de l'entreprise.
2. Identifier les buts de chaque objectif.
3. Identifier les indicateurs de performance.

1.1. Les objectifs principaux

Un bon objectif doit répondre aux contraintes suivantes :

- Mesurable : L'objectif s'exprime en fonction d'une unité de mesure.
- Borné : L'objectif est impérativement défini dans une dimension de temps.
- Réaliste : La Méthode pour l'atteindre est tout à fait possible.
- Accessible : Les moyens sont disponibles, les risques limités.
- Fédérateur : L'objectif recueille l'adhésion de la majorité.
- Constructif : L'objectif contribue à la démarche de progrès.

Sagemcom membre de l'alliance LoRa travaille sur toute la partie radio et le cœur du réseau.

Parmi ses objectifs principaux :

- Assurer l'ajout des nouveaux objets.
- Contrôler la charge du réseau.
- Mesurer la qualité de la communication.

1.1.1. Objectif 1 : Assurer l'ajout de nouveaux objets

Chaque objet doit passer par un processus de jointure avant de commencer à échanger des messages. La communication dans un réseau LoRa se fait entre les objets et les applications à travers les passerelles. Une seule infrastructure peut supporter plusieurs réseaux. Donc, Notre rôle consiste à suivre l'évolution des objets connectés, des applications servies, des réseaux montés et des passerelles déployées. Aussi, nous devons s'assurer du bon déroulement du mécanisme de jointure.

1.1.2. Objectif 2 : Contrôler la charge du réseau

La communication dans un réseau IOT est bidirectionnelle et les messages échangés peuvent être liés aux applications servies. Pour contrôler la charge du réseau il faut compter les messages échangés entre les différents types des messages et superviser la bande passante occupée entre le flux montant et le flux descendant.

1.1.3. Objectif 3 : Mesurer la qualité de la communication

Les échanges sans fil dans notre réseau LoRa peut lever des problèmes liés à la qualité de signal. Ce qui peut provoquer des messages erronés ce qui engendre une perte d'information. Pour cela, Sagemcom tient à superviser les facteurs RSSI et SNR des messages échangés et l'évolution du nombre des messages erronés.

1.2. Les Indicateurs de performance

1.2.1. Nombre de réseaux

Plusieurs clients peuvent exploiter notre infrastructure. Pour séparer le trafic de chaque client, Sagemcom doit donc gérer plusieurs réseaux coexistants. Le nombre des réseaux reflète le nombre des clients dans notre système. Cet indicateur est calculé par une agrégation de comptage sur le champ « NetworkID ». Il est un indicateur de haut niveau d'abstraction.

1.2.2. Nombre d'applications

Le premier rôle de notre infrastructure est de garantir la communication entre les applications et les dispositifs. Le nombre des applications est un indicateur de haut niveau dans notre système d'analyse. Il est calculé par une agrégation de comptage sur le champ « AppEUI ». Il reflète la charge réseau mesurée.

1.2.3. Nombre de passerelles

Le nombre des passerelles nous informe sur la maturité de notre infrastructure. Avec cet indicateur, nous pouvons imaginer la zone de couverture et la qualité de notre réseau. Cet indicateur est calculé par une agrégation de comptage sur le champ « GatewayID ».

1.2.4. Nombre d'objets connectés

Le nombre des dispositifs actifs dans notre réseau est un indicateur majeur dans notre système d'analyse. Il représente le nombre des objets actifs dans notre réseau. Cet indicateur est calculé par une agrégation de comptage sur le champ « DevEUI ». Il reflète la charge réseau utilisée.

1.2.5. Évolution des demandes de jointure

Avant de commencer l'envoi des données aux applications, chaque dispositif doit demander de joindre le réseau. Même les dispositifs inscrits au réseau ré-envoient des messages de jointure pour vérifier leurs états. Cet indicateur est calculé par une agrégation de comptage sur le nombre de messages ayant un « MType » de valeur « Join Request ». Ces messages ne doivent pas prendre un comportement abusif car ils peuvent réduire les performances du réseau.

1.2.6. Évolution d'acceptation de jointure

Chaque demande de jointure conforme est suivie d'une acceptation envoyée depuis le serveur du réseau LoRa vers le dispositif. Il faut donc suivre les acceptations de jointure avec les demandes. Cet indicateur est calculé par une agrégation de comptage sur le nombre de messages ayant un « MType » de valeur « Join Accept ».

1.2.7. Évolution des messages échangés

La communication entre les applications et les dispositifs est bidirectionnelle. Chaque message envoyé peut générer un ou plusieurs documents. Cet indicateur est calculé par une agrégation de comptage sur le nombre des messages. Le suivi d'évolution du nombre des messages dans le temps peut être projeté par type de message (Uplink, Downlink) ou nature de communication (Confirmed Data Up, Unconfirmed Data Up, Confirmed Data Down, Unconfirmed Data Down, Join Request, Join Accept).

1.2.8. Évolution de la bande passante occupée

Chaque trame échangée contient des données protocolaires avec ou sans données applicatives. La taille de ces données varie d'un message à un autre. Ces indicateurs sont calculés par des agrégations de sommation sur les champs « PHYPayloadSize » et « FRMPayloadSize » respectivement. Il nous informe sur la bande passante occupée.

1.2.9. Évolution de la qualité de transmission.

Dans la communication non filaire, la puissance des signaux reçus et le rapport signal sur bruit sont des indicateurs de la qualité de la transmission des messages. Ces indicateurs sont calculés par des agrégations de maximum et de minimum sur les champs « RSSI » et « SNR » respectivement. Nous devons suivre l'évolution de ces indicateurs dans le temps pour connaître l'influence des conditions extérieures sur notre infrastructure.

1.2.10. Évolution des messages erronés

Il arrive parfois de recevoir des messages erronés. Le protocole LoRa est capable de détecter ces anomalies. Le nombre de ces messages est un indicateur de la qualité de transmission. Cet indicateur est calculé par une agrégation de comptage sur le nombre des messages ayant une valeur différente de « NoError » dans le champ « FrameErrorCode ».

2. Les tableaux de bord

L'étape de la réalisation des tableaux de bord doit être précédée par une phase d'étude car chaque dashboard apporte les informations les plus pertinentes pour l'utilisateur qu'il consulte. Nous allons tout d'abord identifier les acteurs et leurs champs de responsabilités. Ensuite, nous allons fixer les besoins en termes de QoS pour chaque utilisateur.

2.1. Identification des acteurs

Les acteurs de notre application de visualisation sont :

- Administrateur : L'administrateur a la visibilité sur tous les réseaux avec leurs différents composants.
- Administrateur réseau : Cet utilisateur à une visibilité sur un ou plusieurs réseaux suivant ses droits d'accès.

- Administrateur de la zone géographique : Il supervise le réseau à travers un ensemble de passerelles qui appartiennent à une zone géographique commune.
- Administrateur d'application : Il suit l'état de connexion des objets qui communiquent avec son application.
- Client : C'est l'utilisateur possesseur des objets connectés.

2.2. Définition des objectifs

Les objectifs de chaque acteur se basent sur son métier. Le tableau, ci-dessous, va nous servir pour la conception des tableaux de bord pertinents et nous permettre de bien présenter les informations à leurs consommateurs.

Tableau 4 : Tableau de définition des objectifs d'analyse

Utilisateur	Objectifs
Administrateur	Assurer du bon déroulement du processus de jointure.
	Visualiser l'état de liaison de chaque objet
	Suivre la charge de tous les réseaux à travers le temps
Administrateur réseau	Suivre la charge de ses réseaux à travers le temps
	Visualiser l'état de liaison des objets qui appartiennent à ses réseaux
	Analyser le flux échangé les objets qui appartiennent à ses réseaux
Administrateur zone géographique.	S'assurer de la qualité de transmission entre les objets et ses passerelles.
	Superviser la charge réseau appliquée à ses passerelles.
	Analyser le flux qui arrive à ses passerelles
Administrateur application	S'assurer de la réception des messages liés à ses applications.
	Superviser la charge du réseau consommée.
	Analyser le trafic réseau de ses applications.
Client	S'assurer de l'envoi des messages par ses objets.
	Superviser la charge du réseau consommée.
	Visualiser la liste des messages envoyés par ses objets.

Conclusion

Nous avons listé dans ce chapitre les différents indicateurs de performance qui permettent la supervision de la qualité de service du réseau LoRa et nous les avons détaillé un par un. De plus, nous avons présenté les différents profils des utilisateurs de notre solution et les objectifs attendus par chaque tableau de bord. Dans le prochain chapitre, nous détaillerons les outils et les langages utilisés ainsi que les interfaces des dashboards réalisés.

CHAPITRE 5 : RÉALISATION

Introduction

Dans ce chapitre nous allons définir d'abord les environnements de développement et de déploiement. Après nous allons s'intéresser à la mise en place du système d'analyse et de la réalisation de l'application de visualisation. À la fin de ce chapitre, nous allons donner le chronogramme prévisionnel des tâches du projet.

1. Environnement de travail

1.1. Environnement matériel du système d'analyse

Le cluster d'analyse de données est déployé sur trois machines pour répondre aux exigences minimales requises, dont leur configuration est décrite ci-dessous :

- PC 1 : Intel i5, 8Go Ram, 500Go Disque dur.
- PC 2 : Intel i5, 4Go Ram, 500Go Disque dur.
- PC 3 : Intel i3, 4Go Ram, 250Go Disque dur.

1.2. Environnement matériel de développement

Durant le stage, nous avons réalisé la partie de recherche et développement de notre projet sur un ordinateur de bureau dont la configuration est décrite ci-dessous :

- Processeur : inter i5
- Ram : 8 Go
- Disque dur : 500 Go

1.3. Environnement logiciel

Tout au long de la phase de réalisation de notre solution, nous avons utilisé une multitude d'outils dont nous spécifions :

- Système d'exploitation : Distribution GNU/Linux Ubuntu 14.04.
- IDE : Netbeans 8.1, Geany.
- Langage de développement : java, python.
- Technologies web : Html, CSS, JavaScript, Ajax, JQuery.
- navigateur : Mozilla Firefox

2. Déploiement du système d'analyse

2.1. Déploiement de la plateforme Hortonworks

La première étape du déploiement de notre solution de calcul des indicateurs de performance consiste à mettre en place le Framework Hadoop. Dans la plateforme Hortonworks nous allons activer les services HDFS, Kafka, Storm et Zookeeper que nous utiliserons pour la solution d'analyse de la QoS les autres seront activées suivant d'autres besoins d'analyses ultérieurs. Le diagramme de déploiement est présenté dans la figure 16.

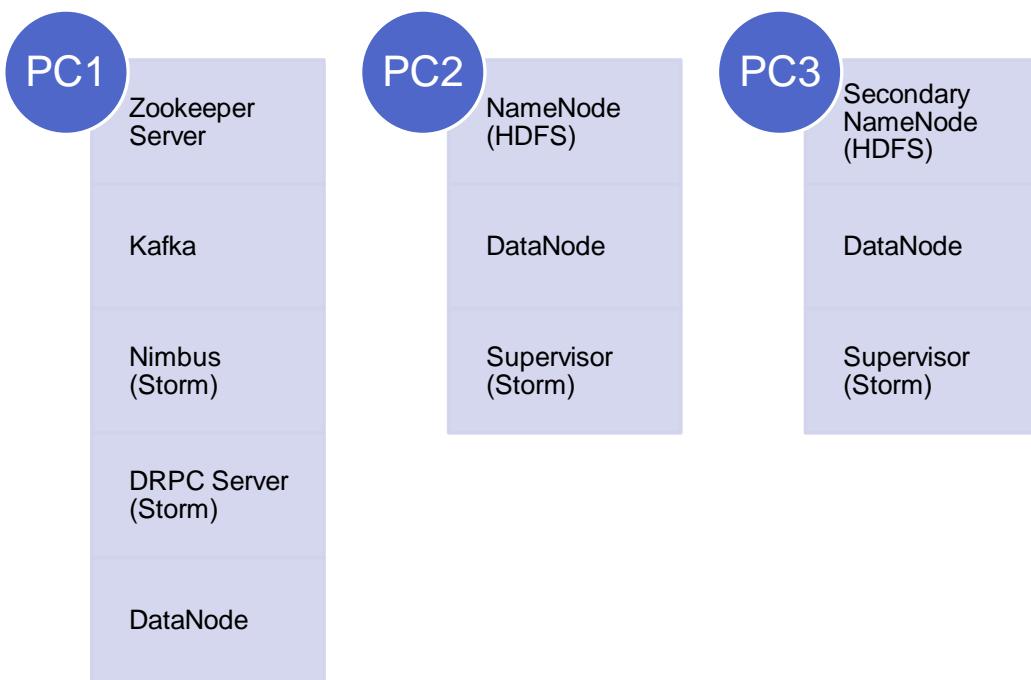


Figure 16 : Architecture du déploiement de Hortonworks

Chaque service fonctionne avec des modules maîtres et des modules clients. Nous avons opté pour une répartition qui permet la meilleure distribution des charges sur les différentes machines.

2.2. Déploiement de l'entrepôt de données Druid

Druid comporte plusieurs modules chacun ayant des rôles différents dans l'architecture. Entre modules maître, de données et des requêtes, nous avons opté pour une séparation lors du déploiement par rôle. Les détails de déploiement sont affichés dans la figure 17.

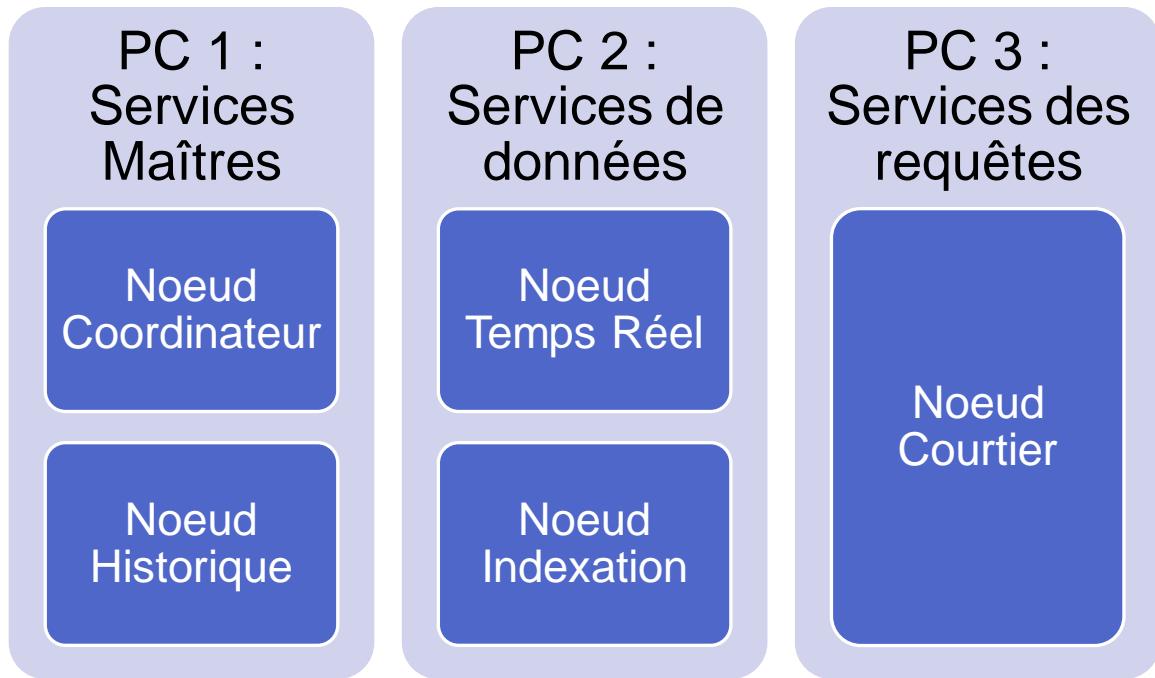


Figure 17 : Architecture de déploiement du cluster Druid

3. Implémentation de l'application de visualisation

Dans ce paragraphe, nous allons naviguer à travers les interfaces les plus pertinentes de l'application de visualisation afin de présenter les travaux réalisés dans ce module.

3.1. Écran d'accueil

Le lancement de l'application commencera par l'interface d'authentification présente dans la figure 18.

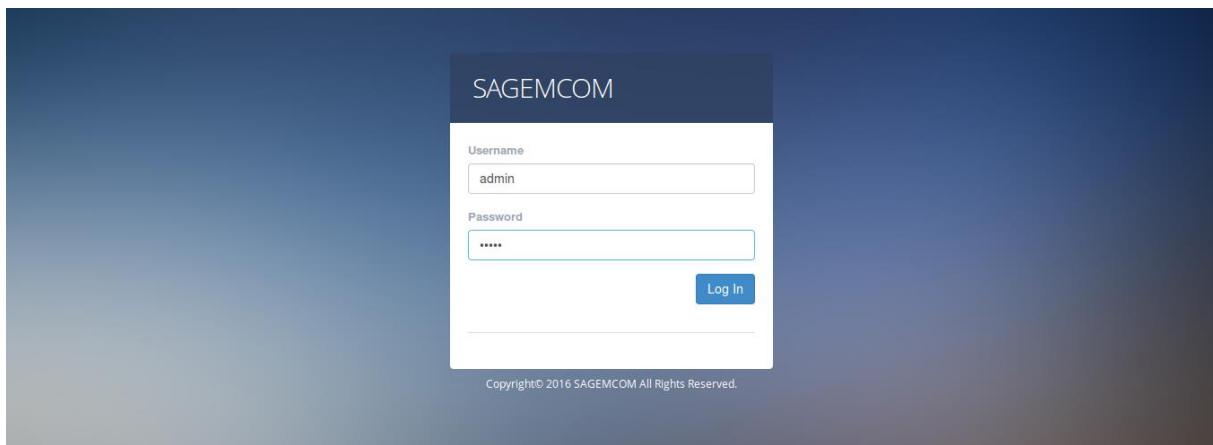


Figure 18 : Interface d'authentification utilisateur

À l'ouverture de l'application, l'utilisateur doit s'authentifier pour pouvoir utiliser l'application. Après l'authentification, chaque utilisateur accède au tableau de bord spécifique à son rôle.

Chaque dashboard est subdivisé en trois zones. La première zone est celle à gauche qui contient le menu de navigation vers les autres interfaces. La zone 2 contient la cartographie des positions des objets et les indicateurs de haut niveau qui sont le nombre de réseaux, le nombre d'objets connectés, le nombre d'applications et le nombre de passerelles. Chaque utilisateur aura une vision sur les indicateurs spécifiques à son champ d'activité. La zone 3 contiendra les graphes qui permettent de répondre aux besoins de chaque utilisateur.

3.2. Interface dashboard « Administrateur »

Une fois l'administrateur est authentifié, son tableau de bord s'affichera comme page de démarrage. Ce dashboard est présenté dans les figures 19 et 20. La partie supérieure du dashboard est consacrée aux positions de géolocalisation des passerelles, des objets et aux indicateurs du haut niveau. Nous trouvons après les graphes suivants :

- **La liste des objets les plus actifs** : C'est un tableau qui contient la liste des cinq objets les plus actifs par rapport au nombre des messages captés sur une période de temps.
- **Liste des applications les plus actives** : C'est un tableau qui contient la liste des cinq applications les plus actives par rapport au nombre de messages reçus sur une période de temps.
- **Liste des passerelles les plus actives** : C'est un tableau qui contient la liste des cinq passerelles les plus actives par rapport au nombre des messages captés sur une période de temps.
- **L'évolution des demandes de jointure** : C'est un graphique en ligne sur une échelle de temps qui présente la proportion entre les demandes de jointures et leurs acceptations.
- **La répartition des types de messages** : C'est un camembert qui montre le pourcentage des messages entre le flux montant et descendant.
- **Portion de messages valides** : c'est un camembert qui montre le pourcentage des messages échangés sans erreurs parmi les autres.

- **L'évolution du nombre de messages** : C'est un histogramme sur une échelle de temps qui présente l'évolution du nombre de messages entre flux montant et descendant.
- **La répartition du débit de communication** : C'est un histogramme qui présente le nombre des messages reçus par rapport au débit de communication.
- **La consommation de la bande passante** : C'est un graphique en ligne sur une échelle de temps qui présente l'évolution des tailles de messages échangés par les objets au cours du temps.

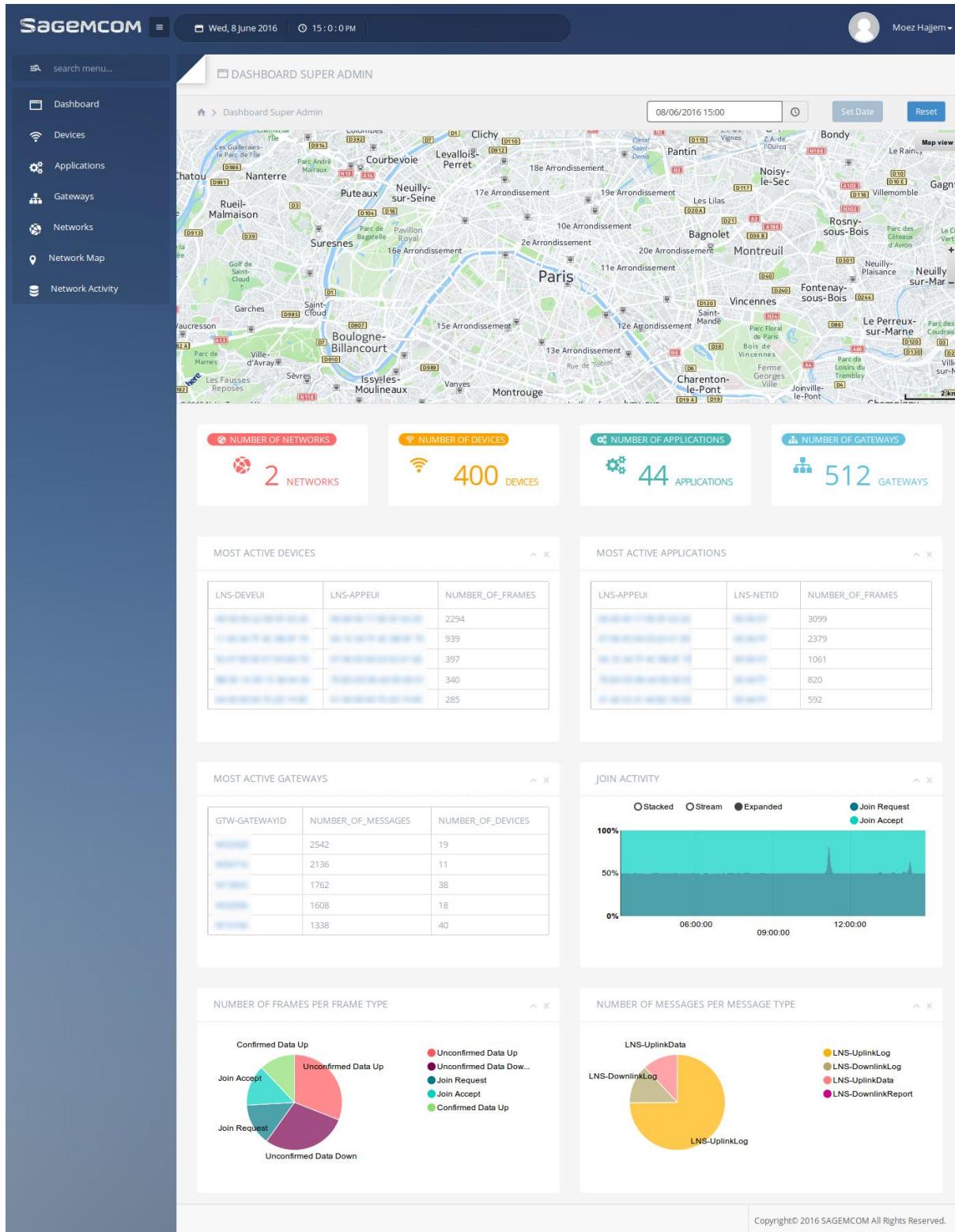


Figure 19 : Interface dashboard administrateur partie 1

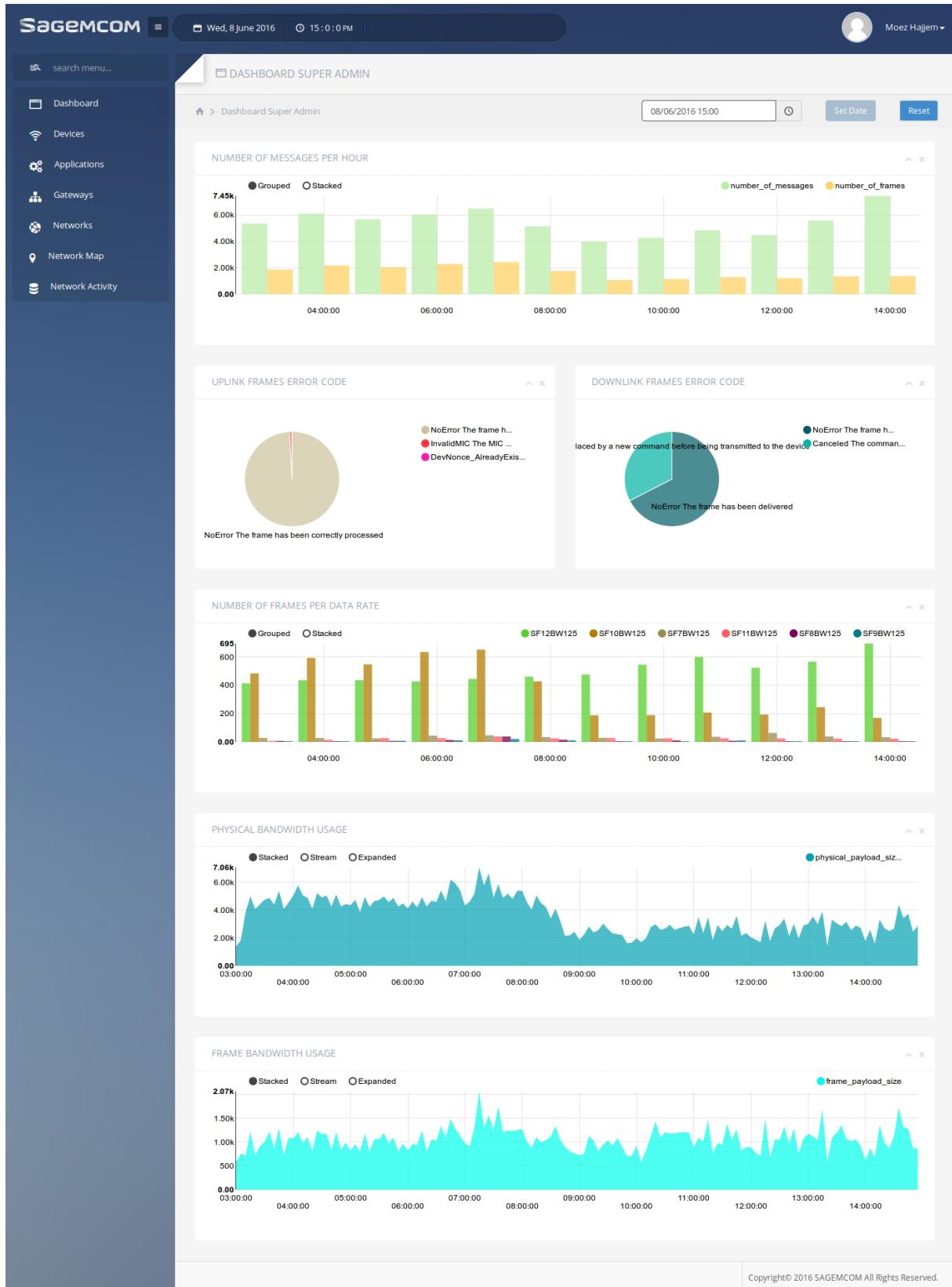
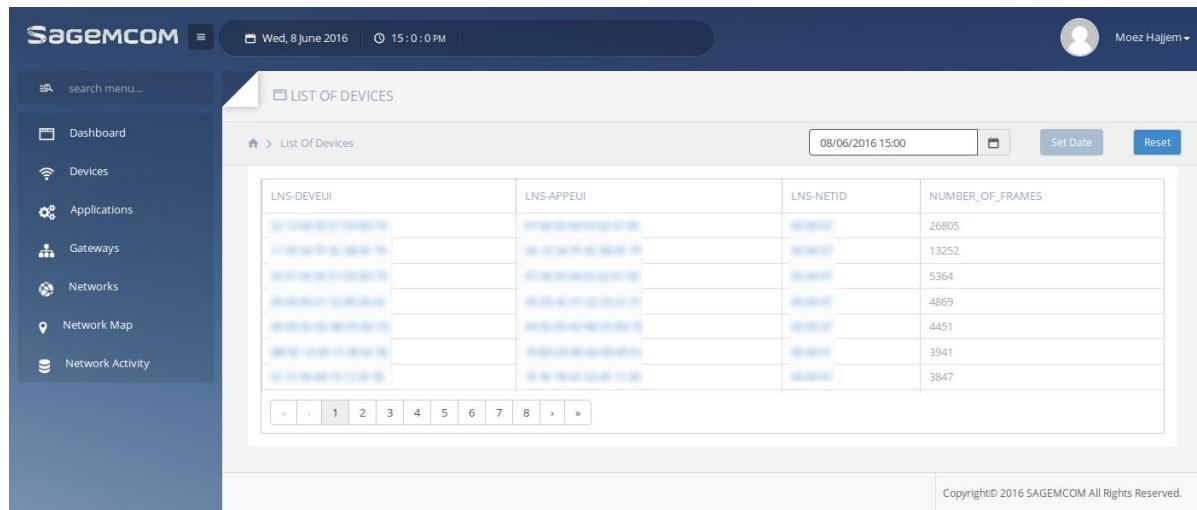


Figure 20 : Interface dashboard administrateur partie 2

3.3. Interfaces des Objets

Le menu « Devices » à gauche permet d'accéder à la liste des objets afin de les parcourir ou pour accéder à leurs tableaux de bord. La figure 21 présente cette liste.



The screenshot shows the Sagemcom web interface. On the left, there's a sidebar with icons for search menu, Dashboard, Devices (selected), Applications, Gateways, Networks, Network Map, and Network Activity. The main area has a header 'LIST OF DEVICES' with a back arrow and a date/time field '08/06/2016 15:00'. Below is a table with four columns: LNS-DEVEUI, LNS-APPEUI, LNS-NETID, and NUMBER_OF_FRAMES. The table contains 10 rows of data. At the bottom is a navigation bar with page numbers 1-8 and arrows.

LNS-DEVEUI	LNS-APPEUI	LNS-NETID	NUMBER_OF_FRAMES
			26805
			13252
			5364
			4869
			4451
			3941
			3847

Figure 21 : Interface de la liste des objets

Le tableau de bord d'un seul objet contient les indicateurs de haut niveau qui sont le nombre des messages et la quantité de données échangées. Nous définissons un onglet pour les graphes (Figure 22), un onglet qui contient le tableau d'activité spécifique à cet objet. Et un onglet qui contient l'historique des positions de cet objet.

Ce dashboard contient les indicateurs de haut niveau comme le nombre de messages, le nombre de frames, la somme des tailles de messages et de frames. En plus les graphes qui permettent de visualiser :

- **La couverture des objets** : c'est un histogramme qui permet de visualiser le nombre de passerelles qui couvrent un objet à chaque envoie de message.
- **La puissance du signal** : c'est un graphique en ligne qui permet de visualiser la puissance du signal reçu "RSSI" et le rapport signal sur bruit "SNR" pour les 20 derniers messages envoyés.
- **La liste des passerelles qui détecte les messages de cet objet.**
- **La consommation de la bande passante.**
- **Portion de messages valides.**

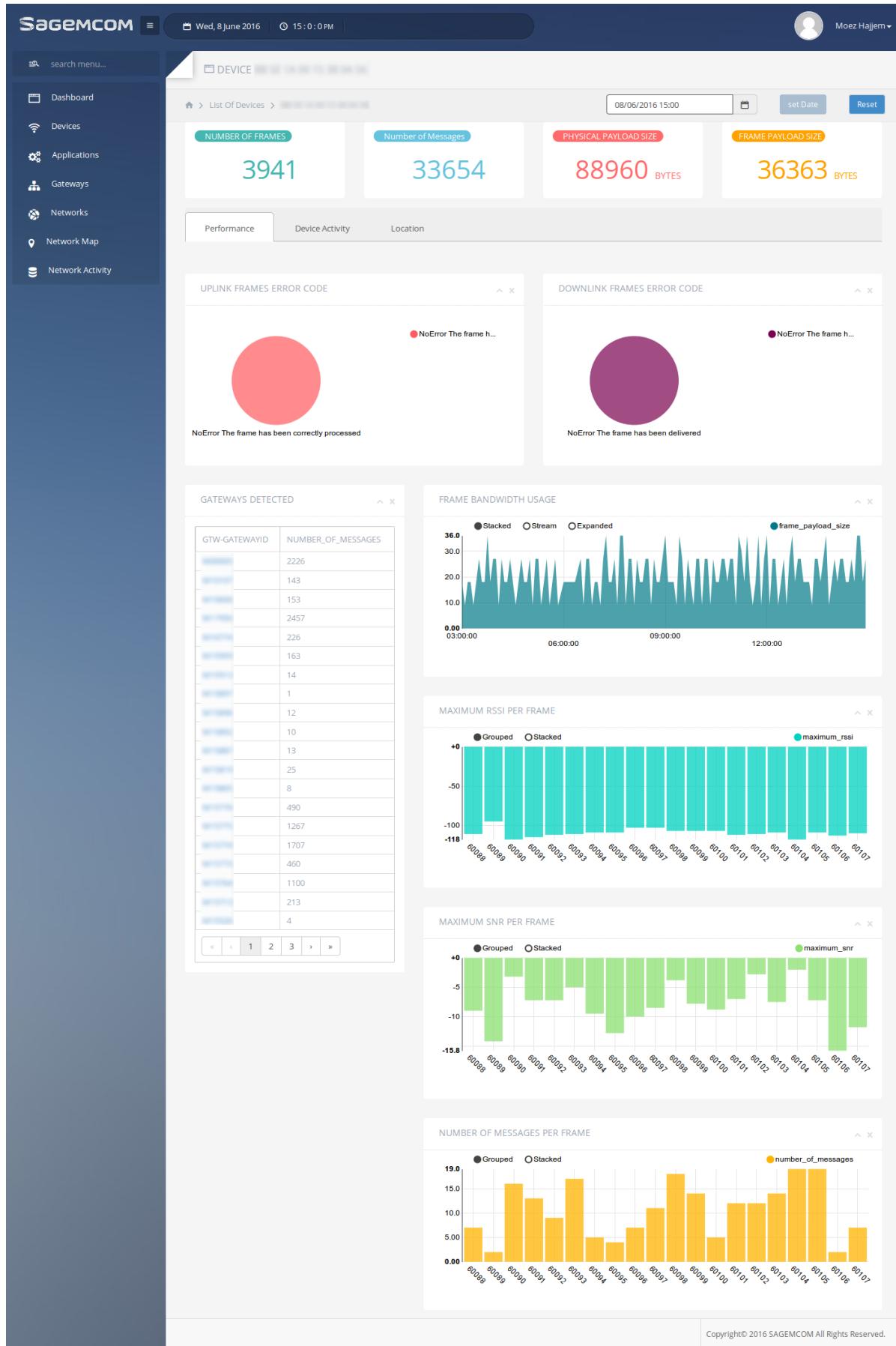
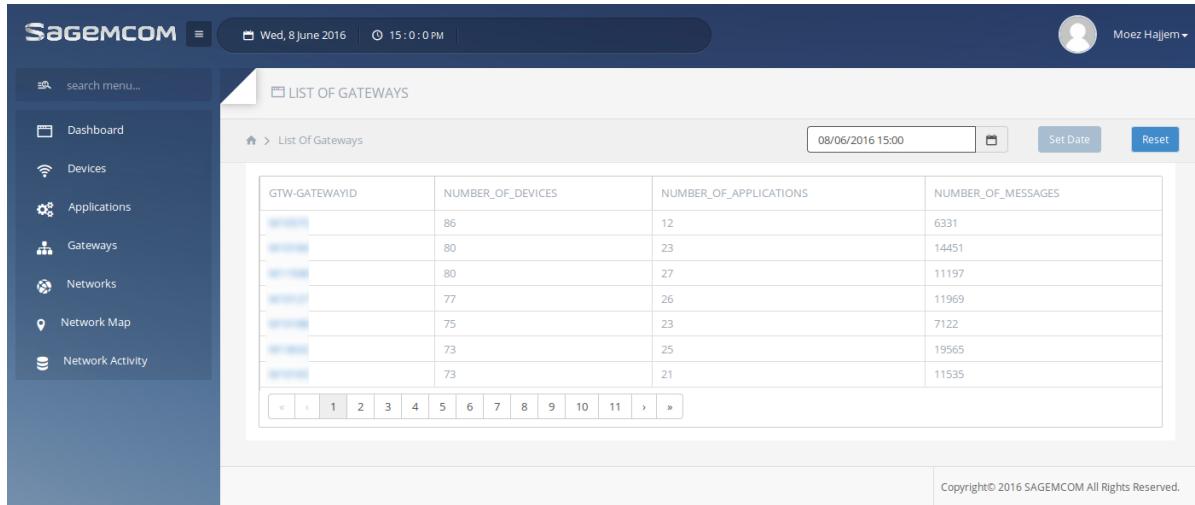


Figure 22 : Interface tableau de bord objet

3.4. Interfaces des applications

L'administrateur peut consulter l'ensemble des applications à travers le menu « Applications » qui permet l'accès à la liste des applications (Figure 23).



The screenshot shows the Sagemcom management interface. On the left is a sidebar with icons for search, dashboard, devices, applications, gateways, networks, network map, and network activity. The main area is titled 'LIST OF GATEWAYS' and shows a table of gateway statistics. The table has columns: GTW-GATEWAYID, NUMBER_OF_DEVICES, NUMBER_OF_APPLICATIONS, and NUMBER_OF_MESSAGES. The data is as follows:

GTW-GATEWAYID	NUMBER_OF_DEVICES	NUMBER_OF_APPLICATIONS	NUMBER_OF_MESSAGES
86	12	6331	
80	23	14451	
80	27	11197	
77	26	11969	
75	23	7122	
73	25	19565	
73	21	11535	

At the bottom of the table is a navigation bar with page numbers from 1 to 11. The top right corner shows the user 'Moez Hajjem'.

Figure 23 : Interface de la liste des applications

Chaque application a un Tableau de bord spécifique pour reporter son état. Ce dashboard contient les indicateurs de haut niveau comme dans le dashboard de l'objet, la liste des objets qu'elle communique avec, la consommation de la bande passante et la qualité de transmission des messages (Figure 24). L'administrateur peut aussi consulter le tableau d'activités des objets avec cette application.

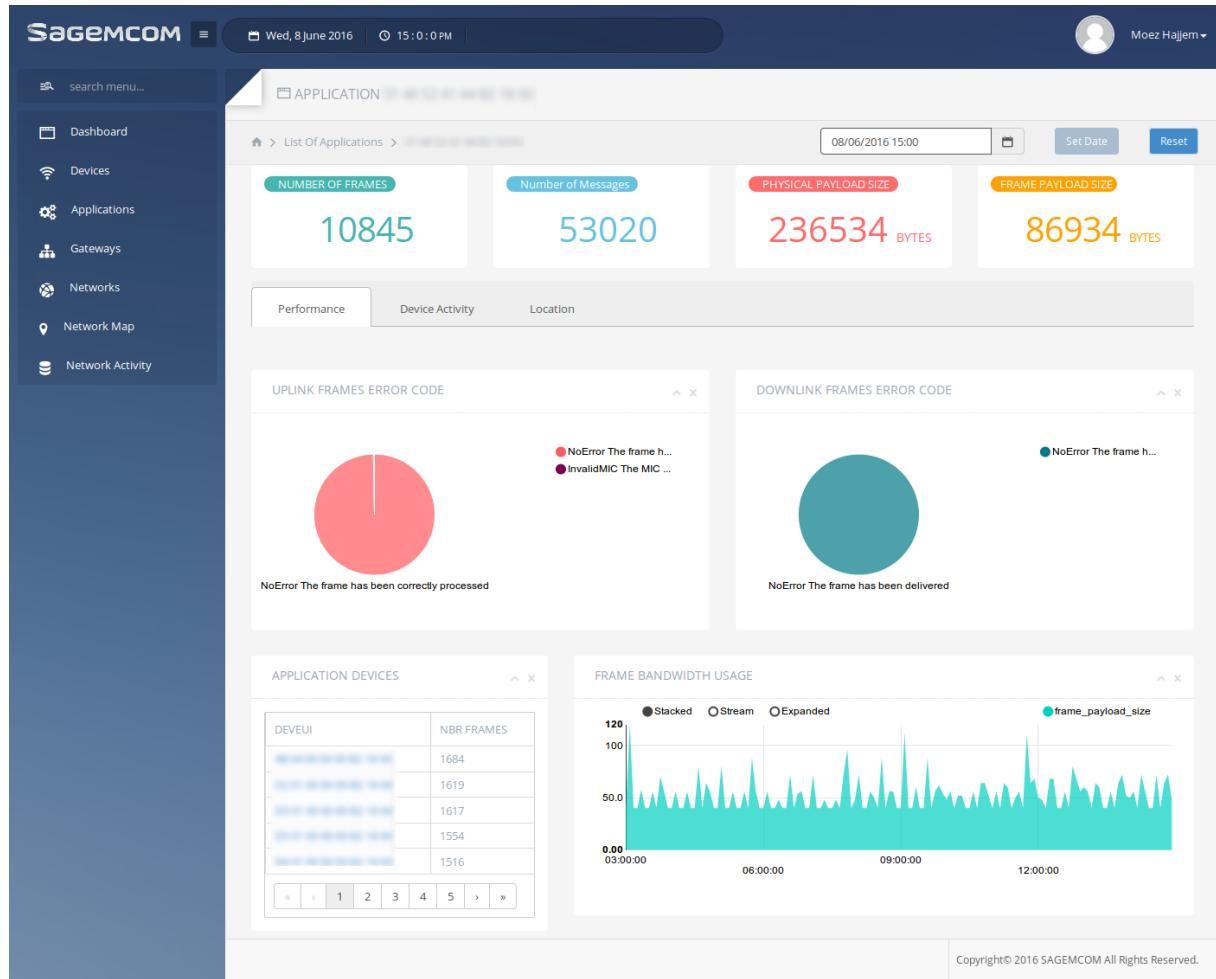


Figure 24 : Interface tableau de bord application

3.5. Interfaces des passerelles

Comme pour les objets et les applications, les passerelles ont une interface pour les lister (Figure 25) et un tableau de bord qui présente leurs états (Figure 26)

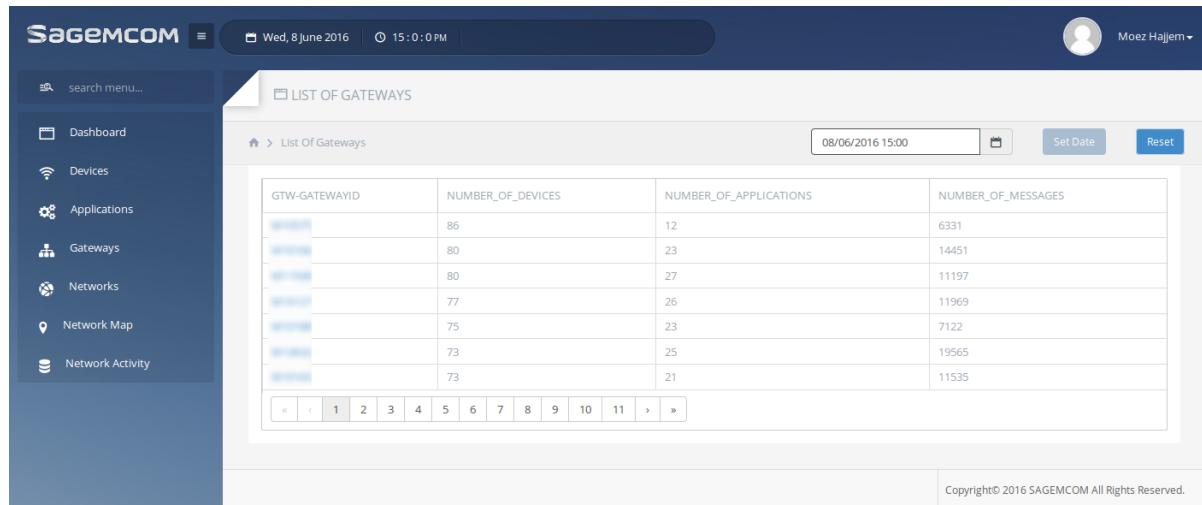


Figure 25 : Interface de la liste des passerelles

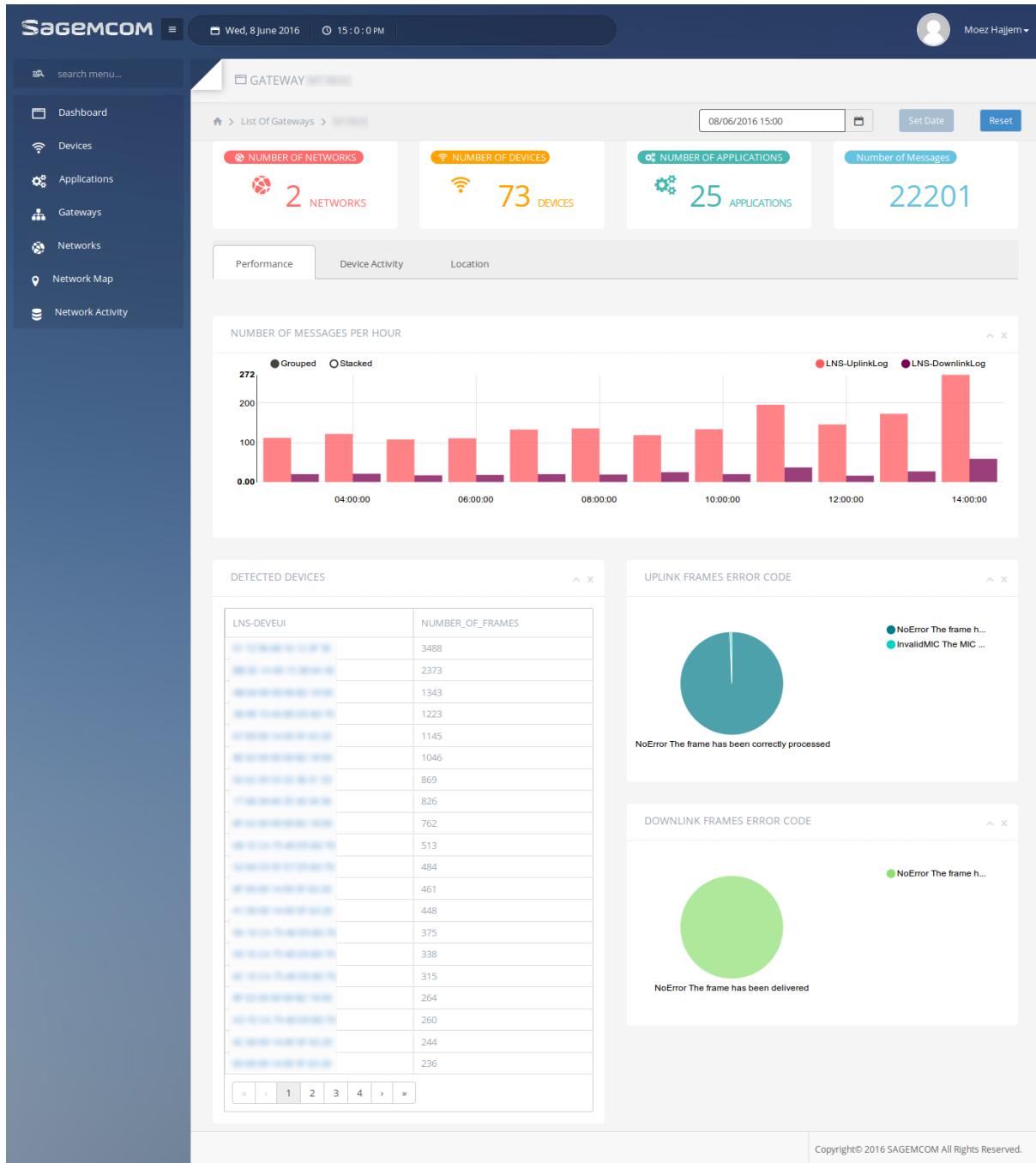


Figure 26 : Interface tableau de bord passerelle

3.6. Interfaces des réseaux

L'ensemble des réseaux sont accessibles depuis le menu « Networks » à gauche (Figure 27). Notre solution est conçue pour supporter plusieurs réseaux de plusieurs clients. Chaque réseau à un tableau de bord propriétaire pour présenter son activité (Figure 28, 29).

The screenshot shows the 'List Of Networks' page. The left sidebar includes a search bar and links to Dashboard, Devices, Applications, Gateways, Networks, Network Map, and Network Activity. The main content area has a header 'LIST OF NETWORKS' and a breadcrumb 'List Of Networks'. It features a table with three columns: LNS-NETID, NUMBER_OF_APPLICATIONS, and NUMBER_OF_DEVICES. Two rows are visible, with the first row showing values 38 and 272 respectively. A date selector shows '08/06/2016 15:00' and buttons for 'Set Date' and 'Reset'. At the bottom right is a copyright notice: 'Copyright© 2016 SAGEMCOM All Rights Reserved.'

Figure 27 : Interface de la liste des réseaux

The screenshot shows the 'Dashboard' page. The left sidebar is identical to Figure 27. The main content area has a header 'NETWORK' and a breadcrumb 'List Of Networks > [redacted]'. It displays four summary cards: 'NUMBER OF NETWORKS' (1 NETWORKS), 'NUMBER OF DEVICES' (272 DEVICES), 'NUMBER OF APPLICATIONS' (38 APPLICATIONS), and 'NUMBER OF GATEWAYS' (508 GATEWAYS). Below these are two tables: 'MOST ACTIVE DEVICES' and 'MOST ACTIVE APPLICATIONS'. The 'MOST ACTIVE DEVICES' table shows LNS-DEVEUI, LNS-APPEUI, and NUMBER_OF_FRAMES. The 'MOST ACTIVE APPLICATIONS' table shows LNS-APPEUI, LNS-NETID, and NUMBER_OF_FRAMES. At the bottom are two pie charts: 'NUMBER OF FRAMES PER FRAME TYPE' and 'NUMBER OF MESSAGES PER MESSAGE TYPE'. The 'NUMBER OF FRAMES PER FRAME TYPE' chart includes categories like 'Confirmed Data Up', 'Unconfirmed Data Up', 'Join Accept', and 'Unconfirmed Data Down'. The 'NUMBER OF MESSAGES PER MESSAGE TYPE' chart includes categories like 'LNS-UplinkData', 'LNS-DownlinkLog', 'LNS-UplinkLog', and 'LNS-DownlinkReport'. Both charts have legends on the right. A copyright notice 'Copyright© 2016 SAGEMCOM All Rights Reserved.' is at the bottom right.

Figure 28 : Interface tableau de bord réseau partie 1

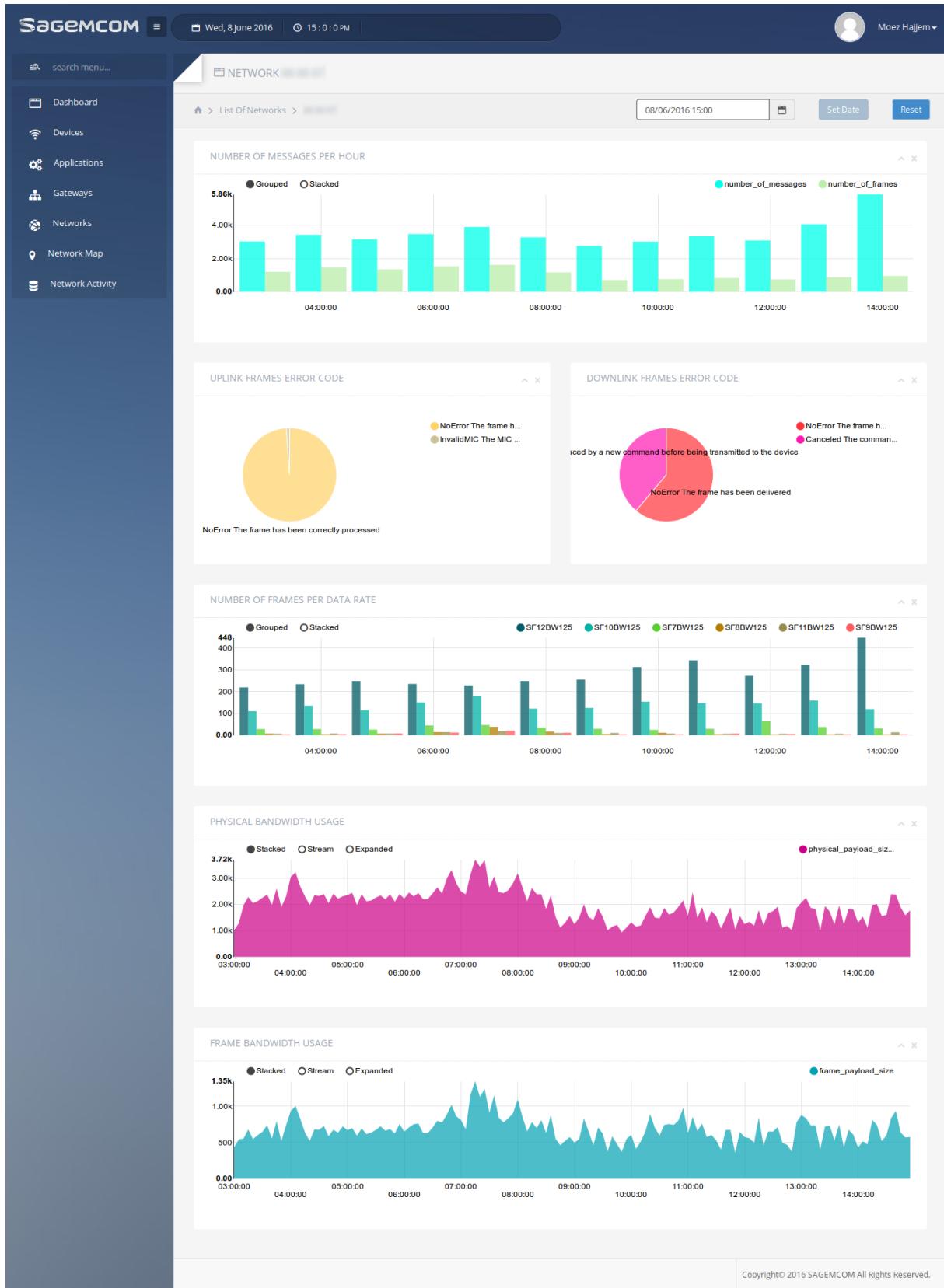


Figure 29 : Interface tableau de bord réseau partie 2

3.7. Interface de l'activité réseau

L'ensemble des frames envoyées par les objets sont affichées en ordre chronologique dans l'interface de l'activité réseau accessible depuis le menu « Network Activity » (Figure 30). Pour chaque Frame, nous avons affiché son temps d'envoie et les identificateurs de l'objet, de l'application et du réseau. Après, nous trouvons le numéro d'ordonnancement, la fréquence, le débit et le « ADR » qui montre si la fonctionnalité d'adaptation du débit est activée. En finalité les indicateurs de puissance du signal, le rapport signal sur bruit et le nombre des messages reçus pour ce frame.

The screenshot shows the Sagemcom Network Activity interface. The left sidebar includes a search bar, a dashboard icon, a devices icon, an applications icon, a gateways icon, a networks icon, a network map icon, and a network activity icon. The main area has a header 'NETWORK ACTIVITY' with a date/time filter set to '08/06/2016 15:00'. Below is a table with columns: TIME, DEVEUI, APPEUI, NETID, FCNT, FREQUENCY, DATARATE, ADR, MAX RSSI, MAX SNR, and NBR MESSAGES. The table lists numerous rows of data corresponding to network frames sent at various times between 2:59:28 PM and 2:56:02 PM. The bottom of the table has navigation links from 1 to 20 and beyond. The footer of the page says 'Copyright© 2016 SAGEMCOM All Rights Reserved.'

TIME	DEVEUI	APPEUI	NETID	FCNT	FREQUENCY	DATARATE	ADR	MAX RSSI	MAX SNR	NBR MESSAGES
2:59:28 PM				3145	868.3	SF12BW125	true	-109	-14	4
2:59:24 PM				5293	868.1	SF12BW125	true	-107	1.8	1
2:59:11 PM				1065	867.1	SF10BW125	true	-112	-1.5	4
2:59:07 PM				1	866.5	SF10BW125	false	-95	9.5	5
2:58:46 PM				70	866.5	SF12BW125	true	-101	7.2	3
2:58:45 PM				39754	867.1	SF12BW125	true	-101	-2.2	17
2:58:41 PM				11805	868.3	SF12BW125	true	-108	1.2	2
2:58:36 PM				260	868.5	SF7BW125	true	-91	9.2	2
2:58:19 PM				3	867.3	SF12BW125	true	-93	11.2	15
2:58:15 PM				60107	868.5	SF12BW125	true	-110	-11.8	7
2:58:14 PM				1	866.5	SF10BW125	false	-101	11.5	6
2:58:12 PM				1064	868.1	SF10BW125	true	-111	-6.8	3
2:57:54 PM				1681	866.5	SF7BW125	true	-99	9	4
2:57:52 PM				1876	868.1	SF12BW125	true	-108	-12.2	1
2:57:45 PM				993	868.1	SF12BW125	true	-107	-6.8	1
2:57:44 PM				46099	866.5	SF12BW125	true	-109	4.2	2
2:57:27 PM				1	866.5	SF10BW125	false	-101	7.8	6
2:57:22 PM				5292	867.3	SF12BW125	true	-107	8.2	4
2:57:19 PM				2	867.1	SF12BW125	true	-101	7.5	13
2:57:12 PM				592	868.1	SF12BW125	true	-94	3.8	5
2:56:46 PM				2	867.3	SF10BW125	false	-106	9.8	6
2:56:40 PM				40064	866.5	SF12BW125	true	-112	6.2	8
2:56:31 PM				1	868.3	SF12BW125	true	-99	6	3
2:56:30 PM				0	867.3	SF12BW125	true	-105	-0.2	41
2:56:26 PM				0	867.1	SF7BW125	true	-97	8.8	5
2:56:23 PM				5291	868.5	SF12BW125	true	-104	2.8	1
2:56:21 PM				253	867.1	SF12BW125	true	-97	2.8	8
2:56:19 PM				40079	867.3	SF12BW125	true	-102	3.5	5
2:56:19 PM				30316	868.3	SF10BW125	true	-103	-5.2	3
2:56:13 PM				252	867.3	SF12BW125	true	-102	5	21
2:56:08 PM				251	868.3	SF12BW125	true	-96	-0.5	1
2:56:02 PM				1630	868.3	SF12BW125	true	-100	-0.8	5

Figure 30 : Interface de l'activité réseau

3.8. Interface dashboard « Administrateur Réseau »

Un administrateur réseau n'a seulement accès qu'aux données des réseaux qu'il supervise. Après l'étape d'authentification l'administrateur réseau accède à son tableau de bord (Figure 31 et 32). En plus des indicateurs de hauts niveaux, ce tableau de bord contient les graphes suivants :

- La liste des objets les plus actifs,
- La liste des applications les plus actives,
- La portion de messages valides,
- La consommation de la bande passante.
- L'évolution du nombre de messages.
- La répartition des types de messages.
- La répartition du débit de communication.

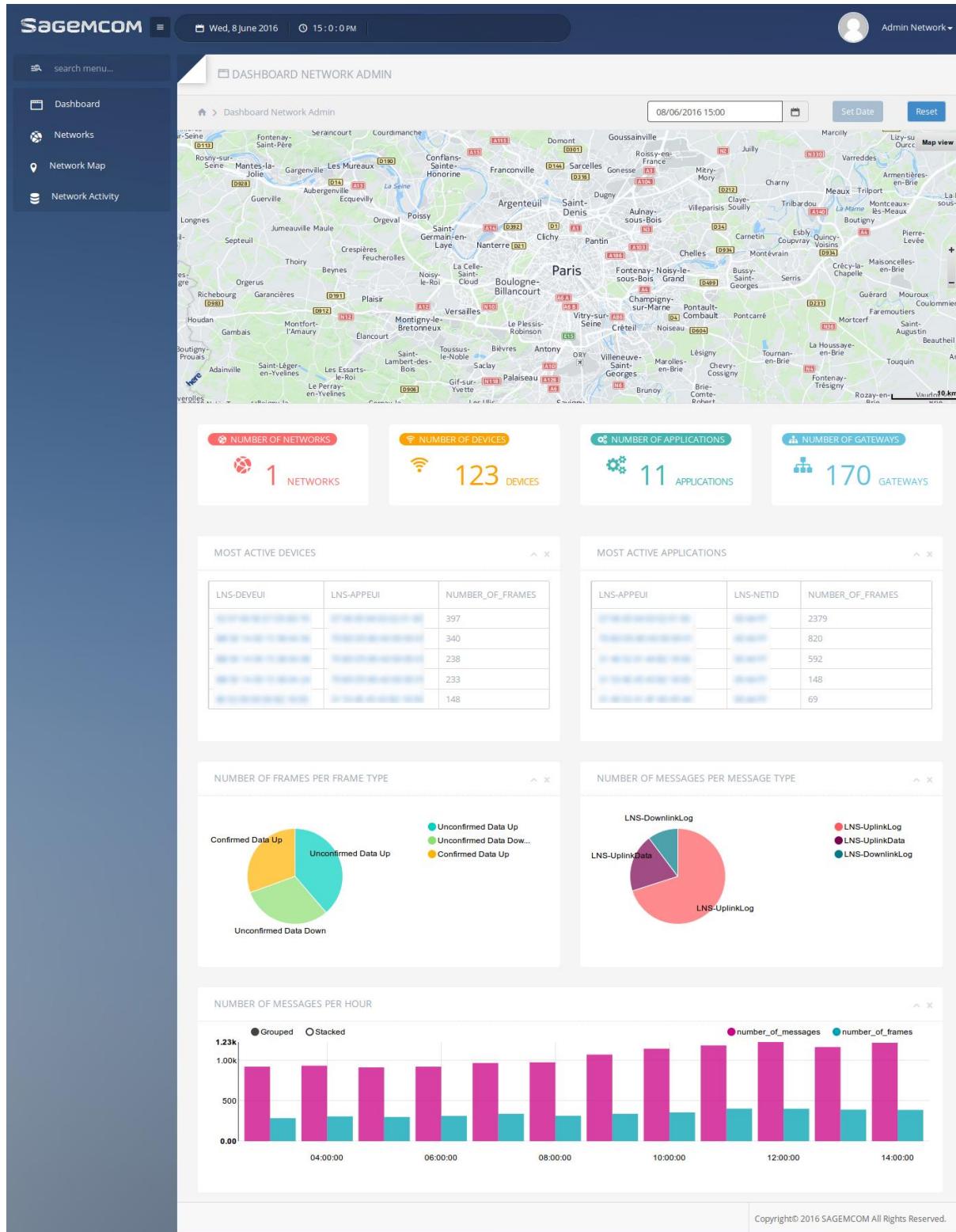


Figure 31 : Interface dashboard administrateur réseau partie 1

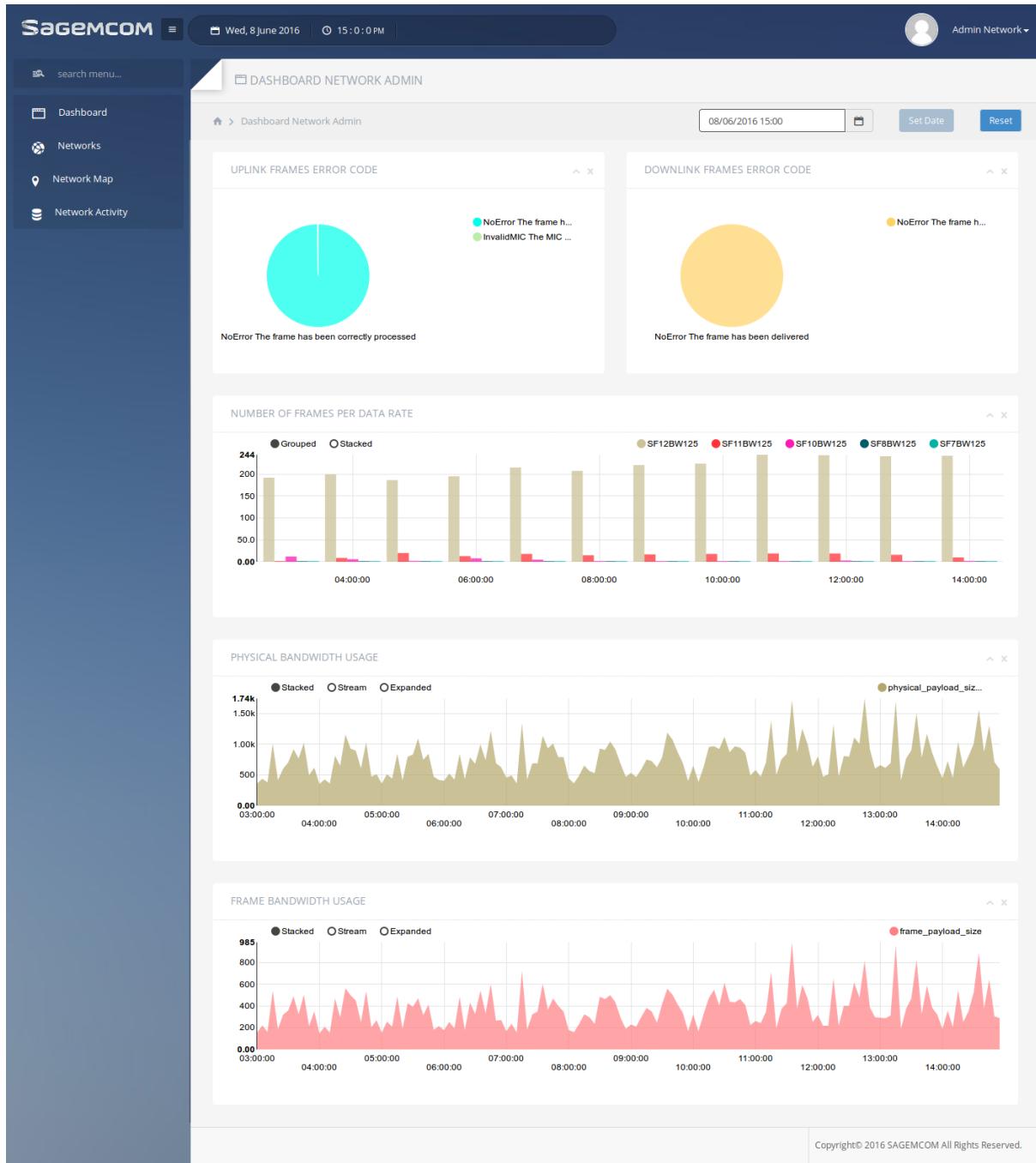


Figure 32 : Interface dashboard administrateur réseau partie 2

3.9. Interface dashboard « Administrateur d'une Zone géographique »

L'administrateur d'une zone géographique est responsable sur un ensemble de passerelles qui appartiennent à une même zone géographique. Une fois connecté, il peut accéder à son tableau de bord présenté par la figure 33. Cet utilisateur aura accès aux interfaces des passerelles suivant ses droits d'accès aux données. Son interface d'accueil contient les graphes :

- La liste des passerelles les plus actives,
- La partie de messages valides,
- L'évolution du nombre de messages,
- La répartition du débit de communication.

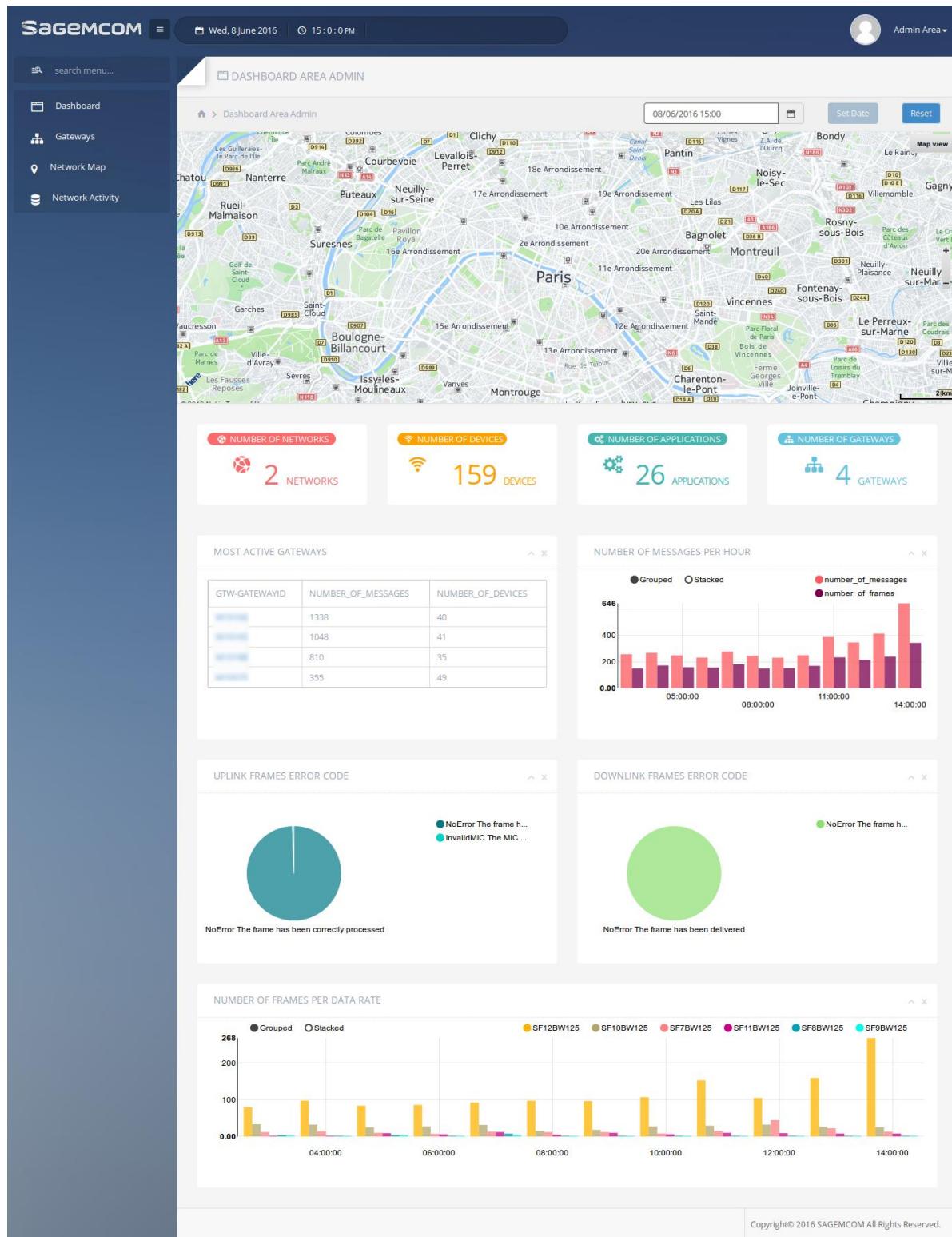


Figure 33 : Interface dashboard « Administrateur d'une zone géographique »

3.10. Interface dashboard « Administrateur Application »

Le tableau de bord de l'administrateur de l'application contient les graphes qui reflètent la charge réseau utilisée par les objets qu'il communique avec (Figure 34). Cet utilisateur peut superviser une ou plusieurs applications, dans ce cas il aura un tableau de bord qui résume toute l'activité de ces applications et un menu pour accéder à l'interface de chaque application à part.

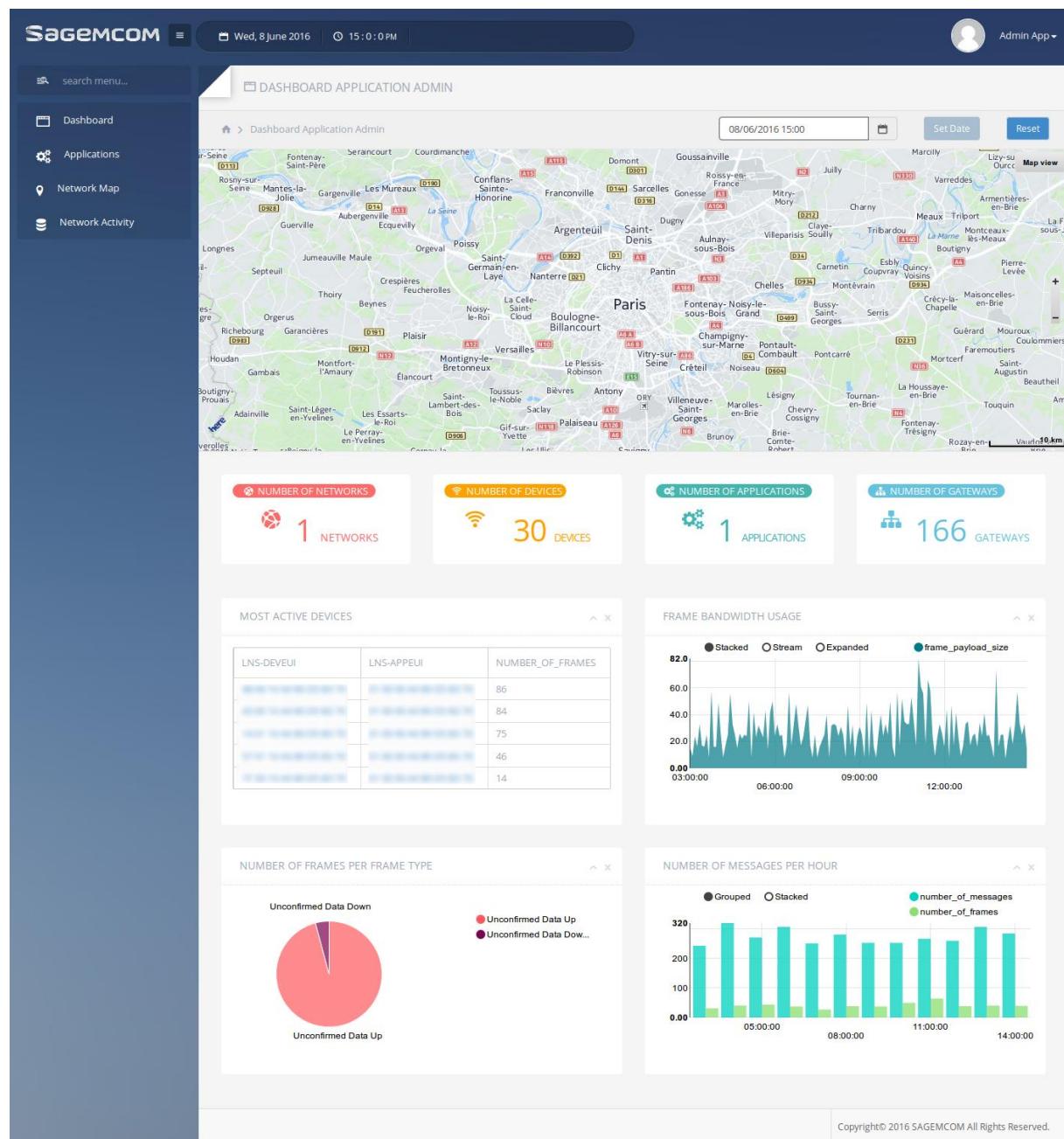


Figure 34 : Interface dashboard « Administrateur application »

3.11. Interface dashboard « Client »

Chaque client a un ensemble d'objets sous sa supervision. Il utilise le tableau de bord de la figure 35 pour suivre leurs états. Il peut consulter aussi les interfaces des objets suivant ses droits d'accès.

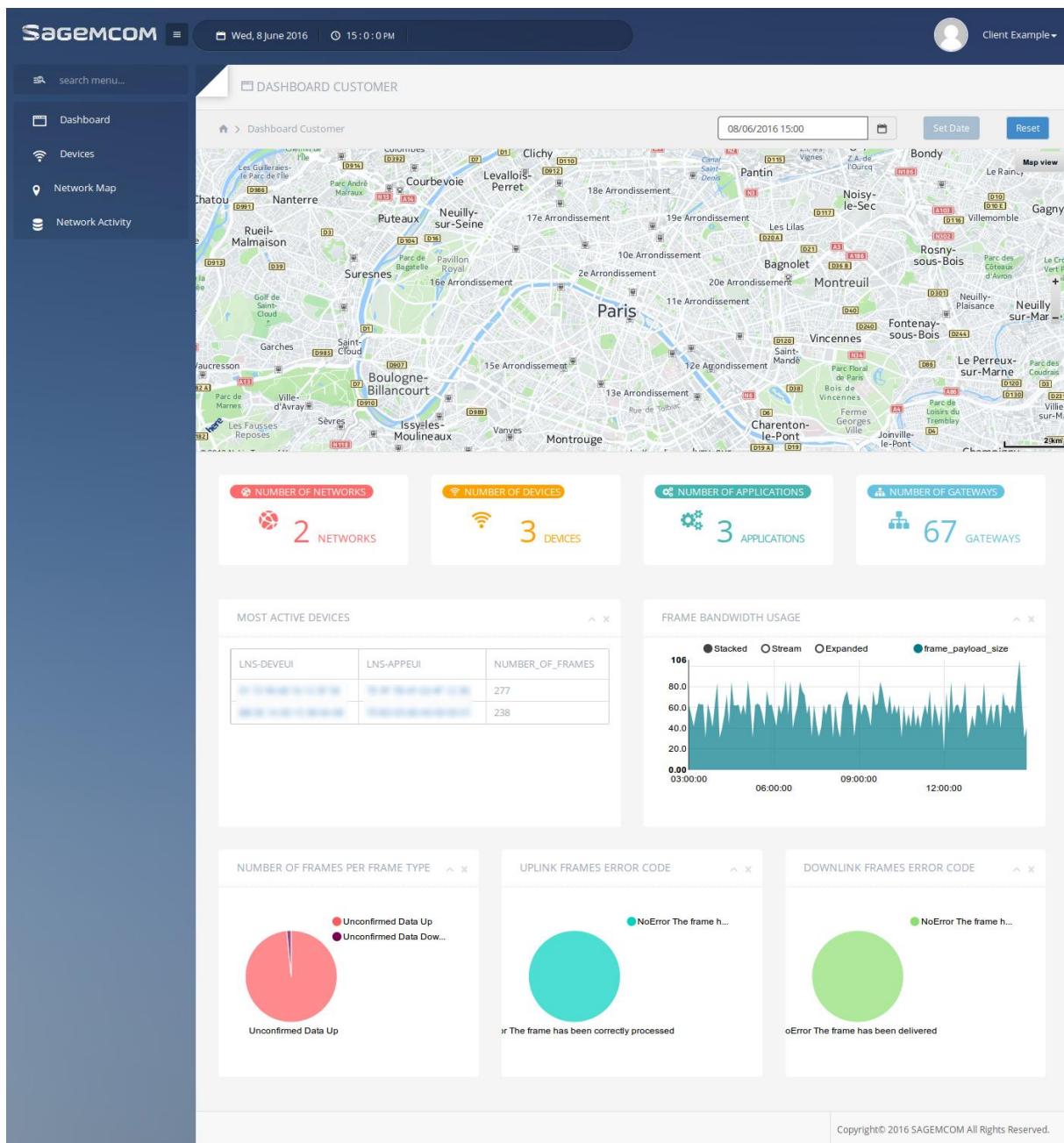


Figure 35 : Interface dashboard du client

4. Chronogramme

Durant les six mois consacrés à ce projet, nous avons fixé des étapes intermédiaires à atteindre pour pouvoir gérer notre temps, garantir de bons résultats et aboutir aux finalités de ce projet.

Nous avons planifié une subdivision en sept étapes :

- ✓ Étape 1 : Se familiariser avec le réseau LoRa et le monde d'internet des objets.
- ✓ Étape 2 : Établir le cadre du projet et établir une spécification des besoins.
- ✓ Étape 3 : Explorer les données et dégager la liste des indicateurs de performance réseau.
- ✓ Étape 4 : Étudier et comparer les solutions permettant d'atteindre nos objectifs.
- ✓ Étape 5 : Mettre en place le système d'analyse de données.
- ✓ Étape 6 : Implémenter les tableaux de bord.
- ✓ Étape 7 : Tester et valider.
- ✓ Rédaction du rapport.

Tableau 5 : Chronogramme de la planification des tâches du projet

Mois	Janvier				Février				Mars					Avril				Mai				Juin						
Semaine	2	3	4	1	2	3	4	1	2	3	4	5	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Étape 1																												
Étape 2																												
Étape 3																												
Étape 4																												
Étape 5																												
Étape 6																												
Étape 7																												
Rédaction du rapport																												

Conclusion

Dans ce chapitre nous avons détaillé la phase de déploiement des différents modules de notre solution sur l'ensemble des nœuds de notre cluster. Puis, nous avons présenté l'application de visualisation et les différents tableaux de bord des utilisateurs à travers des imprimés écran afin de donner une meilleure idée sur le travail réalisé. Et à la fin de ce chapitre, nous avons présenté le chronogramme des tâches.

CONCLUSION ET PERSPECTIVES

L'évolution des technologies de télécommunication a permis l'apparition de nouveaux réseaux destinés à la communication machine to machine et de connecter des objets au réseau Internet. C'est dans ce contexte que s'inscrit notre projet de fin d'études. Ce travail a été effectué au sein de l'équipe anticipation à Sagemcom Software & Technologie. Il s'agit de mettre en place un système de supervision de la qualité de service d'un réseau dédié à l'Internet des objets.

Malgré les contraintes du volume des données et le besoin d'analyse en temps réel, nous avons pu délivrer une preuve de conception fonctionnelle qui répond bien aux objectifs définis au début de ce stage.

En effet ce stage a commencé par l'étude du réseau LoRa et les données échangées par les objets, ce qui nous a permis de dégager les indicateurs de performance clé et pertinentes pour pouvoir mesurer la qualité de service de ce réseau. Ensuite, nous avons consacré une bonne période de temps à concevoir l'architecture optimale qui nous a permis de calculer ces indicateurs en temps réel pour les présenter dans des tableaux de bord.

Notre solution implémentée a été validée avec succès et déployée dans le site de Sagemcom à Paris en vue de passer aux équipes Produit.

Ce stage nous a été très bénéfique non seulement sur le côté technique mais aussi sur le côté humain et professionnel. Il nous a permis d'avoir une première expérience sur les systèmes de traitement massif de données et de consolider nos connaissances d'analyse de données.

Par ailleurs, comme perspective, nous pouvons penser à intégrer dans un premier lieu un module d'alerte qui permet de notifier les administrateurs des cas critiques. Ainsi qu'à ajouter un système de prédiction d'indicateur pour aider les administrateurs dans leurs tâches préventives de maintenance.

WEBOGRAPHIE

- [1] : Site web : Sagemcom (<http://www.sagemcom.com/fr/>, Février 2016)
- [2] : Document : Présentation intégration des stagiaires SST (Janvier 2016)
- [3] : Livre : L'Internet des objets Quels enjeux pour l'Europe (<http://books.openedition.org/>, Février 2016)
- [4] : Article : Tout savoir sur le réseau LoRa de Bouygues (<http://www.objetconnecte.com>), Avril 2016)
- [5] : Présentation : The LoRa-Alliance (Thierry Lestable, Juin 2016)
- [6] : Magazine : Sagemcom COMIN (N°34 – Mars 2015)
- [7] : Article : Qu'est-ce que le Big Data ? (<http://www.redsen-consulting.com/fr>, Mars 2016)
- [8] : Article : Qu'est-ce que le Big Data ? (<http://www.lebigdata.fr>, Mars 2016)
- [9] : Article : Big Data : Panorama des solutions 2016 (<http://blog.ippon.fr/>, Avril 2016)
- [10] : Article : NOSQL base de données « Not Only SQL » (www.lemagit.fr/, Avril 2016)
- [11] : Article : Big Data : La jungle des différentes distributions open source Hadoop (<http://blog.ippon.fr/>, Avril 2016)
- [12] : Site web : Druid (<http://druid.io/>, Avril 2016)
- [13] : Article : Le moteur d'analyse distribuée Kylin d'eBay (<http://www.open-source-guide.com>, Avril 2016)
- [14] : Site web : Apache Kafka (<https://kafka.apache.org/>, Avril 2016)
- [15] : Article : Traitements Big Data avec Apache Spark (<https://www.infoq.com>, Avril 2016)
- [16] : Site web : Apache Storm (<http://storm.apache.org/>, Avril 2016)