

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique
Université de 8 Mai 1945 – Guelma -
Faculté des Mathématiques, d'Informatique et des Sciences de la matière
Département d'Informatique



Mémoire de Fin d'études Master

Filière : Informatique

Option : Science et technologie de l'information et de la communication

Thème :

**Vers une nouvelle méthode de stockage de données
« Big Data » dans un environnement smart city**

Encadré Par :
Dr. N. Benhamida

Présenté par :
Menasria Amar

Juin 2022

Remerciements

Avant tous, je remercie le *Dieu*, de m'avoir aidé à réaliser ce projet.

Mes remerciements vont tout d'abord au corps professoral et administratif de la faculté des Mathématiques, de l'Informatique et des Sciences de la matière, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation qualifiante et perfectionnée.

Nos remerciements vont à notre encadrante **Dr. BENHAMIDA NADJETTE**, pour toute sa gentillesse, pour ses précieux conseils et pour sa patience, ainsi tous ceux qui m'ont aidés et soutenus dans notre travail.

Je remercie également les membres du jury d'avoir accepté d'examiner ce mémoire et leurs remarques précieuses.

Finalement, je tiens à exprimer ma profonde gratitude à ma famille qui m'a toujours soutenue et encouragée.

Table des matières

RESUME.....	8
ABSTRACT.....	9
ملخص	10
INTRODUCTION GENERALE	11
CHAPITRE I : BIG DATA	12
1. Introduction	12
2. Définition du Big Data	12
3. Historique	13
4. Structure de Big Data	14
1. Données structurées.....	14
2. Données non structurées.....	14
3. Données semi structurées	15
4. Données quasi structurées	15
5. Caractéristique Big Data	16
1. Volume	16
2. Vitesse.....	16
3. Variété	16
4. Véracité.....	16
5. Valeur	17
6. Clés des Big Data	17
1. Les données publiques	17
2. Les données privées	18
3. Les échappements des données	18
4. Les données communautaires	18
5. Les données d'auto-quantification.....	18
7. Processus de traitement des Big Data.....	18
1. Génération de données	18
2. Acquisition	19
3. Prétraitement des données	19
4. Stockage.....	21
5. Analyse des données Big Data.....	23
8. Défis du Big Data.....	24
1. Représentation des données :	24
2. Réduction de la redondance et compression des données.....	24

3.	Mécanisme analytique.....	25
4.	Confidentialité des données	25
5.	Gestion de l'énergie.....	25
6.	Extensibilité et évolutivité	25
7.	Coopération	26
9.	Relation entre les méga données et l'IoT	26
10.	Conclusion.....	27
CHAPITRE II : SMART CITY		28
1.	Introduction	28
2.	Définitions.....	28
3.	Historique	29
4.	Les objectifs de la Smart city.....	30
5.	Les outils de la Smart city	31
1.	La technologie	31
2.	Numérique	31
3.	Internet of Things	32
4.	Open Data	32
6.	Les acteurs de la Smart city	33
7.	Les domaines de la Smart city.....	33
1.	Smart Economy	34
2.	Smart Mobility.....	35
3.	Smart Environment.....	35
4.	Smart People	36
5.	Smart Living.....	36
6.	Smart Governance	36
8.	Le Big Data dans un Smart city.....	37
1.	L'administration locale préventive	38
2.	La gestion des opérations locales.....	38
3.	Le développement des réseaux locaux.....	38
4.	La diffusion de l'information locale	39
9.	Défis liés à l'utilisation des Big Data dans les Smart city	39
1.	Gestion de la qualité.....	39
2.	Intégration des données.....	39
3.	Confidentialité des données	40
4.	Comprendre les besoins	40
5.	Valorisation de l'information géographique.....	40

6.	La conception des services des Smart cities	40
10.	Utilisation de TIC dans les Smart city.....	41
11.	Conclusion.....	42
CHAPITREIII : PROPOSITION ET IMPLEMENTATION D'UN NOUVEAU MODELE DE STOCKAGE DE BIG DATA		43
1.	Introduction	43
2.	Le stockage de données dans les villes intelligentes.....	43
1.	Solutions existantes.....	44
2.	Discussion	44
3.	Un algorithme de stockage distribué des données.....	45
1.	Description de l'algorithme proposé.....	46
2.	Les avantages de la solution proposée	48
4.	Exemple du modèle proposé	48
5.	Environnement de développement	49
1.	Représentation de CupCarbon	49
2.	La carte	50
3.	La barre de menu	51
4.	La barre d'outils.....	53
5.	Le menu des paramètres	54
6.	La barre d'état.....	55
7.	La console	55
8.	CupCarbon Objects.....	55
9.	Les langages utilisés.....	56
10.	Python.....	56
11.	Senscript	56
6.	Interface et exécution	57
7.	Conclusion.....	60
CONCLUSION GENERALE.....		61
REFERENCE		62

Table des figures

Figure 1: Volume annuel des données numériques créées à l'échelle mondiale depuis 2010[6].	13
Figure 2 : Structure des données massives [12].	15
Figure 3: Les 5V des Big Data [17].	17
Figure 4: Relation entre Big Data et internet des objets [29].	26
Figure 5: Smart city [37].	30
Figure 6: Les Domaine de la Smart city [55].	34
Figure 7 : Classification des cas d'utilisation du Big Data dans les villes intelligentes [56].	38
Figure 8 : modèle de stockage de données distribué.	46
Figure 9 : Le modele de stockage distubié de Big Data.	48
Figure 10 : Interface utilisateur de CupCarbon [81].	50
Figure 11 : La barre menu de CupCarbon [81].	51
Figure 12: La barre d'outils de CupCarbon [81].	54
Figure 13 : Lancement du modèle implémenté.	57
Figure 14 : génération des données par les capteurs et les envoyer vers les routeurs.	58
Figure 15 : l'envoi de données par les capteurs vers les récepteurs appropriés au travers les routeurs.	58
Figure 16 : redirection des requêtes des capteurs vers un récepteur distant pour le stockage distribué (1).	59
Figure 17 : redirection des requêtes des capteurs vers un récepteur distant pour le stockage distribué (1).	59
Figure 18 : libération de la mémoire et enregistrement des données sur l'espace interne.	60

Table des tableaux

Tableau 1 : Abréviations utilisées pour définir les nœuds IoT	49
Tableau 2 : Les différents arrière-plans de la carte du simulateur CupCarbon.	51
Tableau 3 : La partie barre de menu du simulateur CupCarbon.	53
Tableau 4 : La partie menu des paramètres du simulateur CupCarbon.....	55

RESUME

Ce travail s'inscrit dans le cadre d'un projet de fin de cycle master visant un nouveau modèle de stockage de Big Data dans un environnement Smart City. L'objectif est de proposer un nouveau modèle de stockage distribué (décentralisé) des données massives proviennent de plusieurs sources de la ville intelligente. Ce modèle prend en charge le stockage des données au niveau du périphérique, afin de réduire et d'éliminer plusieurs quantités tels que : le temps de traitement, le temps de réponse réel, et la latence. Dans le cas où la mémoire du périphérique est pleine ou saturée, le model proposé permet d'exploiter les ressources du réseau en assurant la distribution des données sur d'autres appareils afin d'éviter de les avoir perdues. Pour la conception, nous avons opté pour le simulateur CupCarbon et le langage de programmation Python.

Mots clés:Big Data, Smart city, Stockage distribué de données.

ABSTRACT

This work is part of an end of cycle master project aiming at a new model of Big Data storage in a Smart City environment. The objective is to propose a new distributed storage model for massive data coming from several sources in the Smart City. This model supports data storage at the device level, in order to reduce and eliminate several quantities such as: processing time, real response time, and latency. In case the memory of the device is full or saturated, the proposed model allows to exploit the network resources by ensuring the data distribution on other devices in order to avoid losing them. For the design, we opted for the CupCarbon simulator and the Python programming language.

Keywords: Big Data, Smart city, Data distributed storage.

ملخص

هذا العمل هو جزء من مشروع تخرج يهدف إلى اقتراح نموذج جديد لتخزين البيانات الضخمة في المدينة الذكية. الهدف هو اقتراح نموذج تخزين موزع (لامركزي) للبيانات الضخمة القادمة من عدة مصادر. هذا النموذج يدعم تخزين البيانات على مستوى جهاز المستخدم لتحسين كميات المتعددة والقضاء عليها مثل: وقت المعالجة ووقت الاستجابة الفعلي ووقت الانتظار. في حالة امتلاء الذاكرة الطرفية أو تشبعها، يتيح النموذج المقترح استغلال موارد الشبكة من خلال ضمان توزيع البيانات على الأجهزة الأخرى لتجنب فقدها. بالنسبة للتصميم، اخترنا محاكي CupCarbon ولغة برمجة بايثون.

الكلمات المفتاحية: البيانات الضخمة، المدينة الذكية، تخزين البيانات الموزع.

INTRODUCTION GENERALE

Les villes d'aujourd'hui doivent faire face à des problèmes tels que la gestion des ressources, la pollution de l'air, la gestion des déchets, les embouteillages et le vieillissement des infrastructures. Elles doivent également collecter et analyser de grandes quantités de données sur la vie urbaine provenant de nombreuses sources afin de créer un contenu utile pour les citoyens, les visiteurs, les gouvernements et les entreprises.

La révolution de l'information a conquis tous les domaines et a donné naissance à une nouvelle technologie appelée "Big Data". Cette dernière présente un nouveau défi pour les différents chercheurs qui souhaitent gérer et stocker convenablement tous ces énormes ensembles de données.

Le « Big Data » peut permettre de réduire beaucoup de problème qui ont apparait avec le développement des Smart city en incorporant des capteurs de divers utilités. Les données obtenues peuvent être traitées et analysées pour obtenir des informations précieuses et utiles. Aussi, le Big Data peut servir à réduire et résoudre de nombreux problèmes liés au développement des villes intelligentes.

L'objectif principal de ce travail est d'étudier les technologies et les techniques du Big Data utilisés dans un environnement intelligent afin de proposer une nouvelle méthode de stockage distribué de « Big Data » dans un environnement Smart city.

Ce mémoire est divisé en trois chapitres comme suit :

- Dans le premier chapitre, on a commencé par une introduction et des définitions du « Big Data ». Ensuite, on a détaillé la structure des données ainsi que les caractéristiques de ce type de données. Puis, On a présenté le processus du traitement, les défis des « Big Data » ainsi que leur relation avec l'internet des objets (IoT). Enfin, nous concluons le chapitre par une conclusion.
- Dans le deuxième chapitre, on a présenté quelques généralités liées aux Smart cities ou bien ce qu'on appelle en français les villes intelligentes ; passant par une brève définition de ce type de villes, leurs principaux objectifs ainsi leurs outils. En citant aussi les acteurs, les domaines et les différents défis liés à l'utilisation des « Big Data » dans les Smart city.
- Enfin, le troisième chapitre présente la problématique, la solution proposée et les avantages de cette dernière, ainsi qu'une description détaillée du modèle proposé. Enfin, ce chapitre présente les différentes interfaces de notre système ainsi que les outils utilisés dans leurs réalisations (simulateur CupCarbon).

CHAPITRE I : BIG DATA

1. Introduction

Avec l'évolution de la technologie, l'utilisation des données a augmenté rapidement ces dernières années, et dans tous les domaines. Cela a été la cause principale de se retrouver face à de grande quantité de données appelées aujourd'hui Big Data, données massives en français ou encore Méga données. Par conséquence, les chercheurs ont été dans l'obligation de chercher de nouveaux concepts et de nouvelles méthodes de traitement afin de faciliter (1) le stockage, (2) la recherche, (3) le partage, (4) l'analyse et (5) la visualisation de ce type de données pour en extraire des informations pertinentes qui peuvent être utiles pour la prise de décision dans un ou plusieurs systèmes d'information.

Dans ce chapitre, nous allons présenter les concepts associés au domaine du "Big Data".

2. Définition du Big Data

Littéralement, le terme Big Data désigne les "données massives" ou les "méga données". C'est un terme anglais utilisé pour désigner de grands ensembles de données qui ne permettent pas l'utilisation d'outils traditionnels de gestion de bases de données parce qu'il est difficile et dans plusieurs autres cas impossibles de traiter, stocker, analyser et gérer ces données avec les anciennes méthodes [1].

Plusieurs définitions ont été données à la notion de "Big Data". Cependant, aucune n'a été définitivement adoptée car ce concept est assez complexe et sa définition diffère selon les utilisateurs et les fournisseurs de services qui s'y intéressent [1].

Parmi les, nous citons les définitions suivantes :

- Les Big Data sont des ressources d'information volumineuses, rapides et diverses qui nécessitent un traitement créatif et efficace des données pour améliorer la compréhension et la prise de décision [2].
- Contrairement aux données traditionnelles, le terme Big Data fait référence à une croissance jeux de données comprenant des formats hétérogènes : données structurées, non structurées et semi-structurées. Le Big Data est un complexe nature qui nécessitent

des technologies puissantes et des algorithmes plus avancés. Ainsi, les outils traditionnels ne peuvent pas être plus efficaces dans le cas des applications Big Data [3].

- Les méga données désignent les informations dynamiques générées dans des systèmes complexes. Elles peuvent aider les entreprises à révéler des informations cachées et à obtenir des avantages concurrentiels. Il est devenu un domaine d'étude important et a attiré énormément d'attention de la part de l'industrie, du milieu universitaire et du gouvernement [4].

3. Historique

Bien que le terme Big Data est apparu au début des années 1990, il n'est pas un domaine totalement nouveau. Au fait, les chercheurs ont beaucoup essayé, au cours de nombreuses années, d'utiliser des techniques d'analyse et de traitement des données pour faciliter leur processus de prise de décision [5].

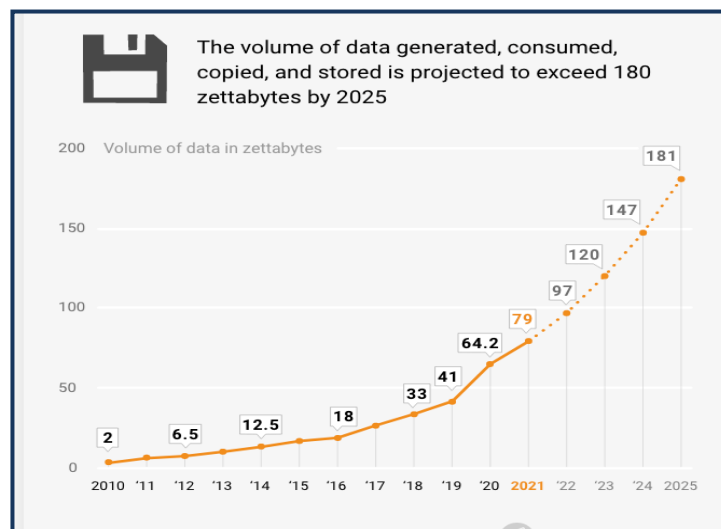


Figure 1: Volume annuel des données numériques créées à l'échelle mondiale depuis 2010[6].

Cependant, le volume global de données dans le monde explose, passant de 6,5 zettaoctets en 2012 à 79 zettaoctets en 2021 (figure 1). Sachant que 79 zettaoctets est équivalent à 79 trillions de giga-octets. Même avec les outils modernes d'aujourd'hui, l'analyse de toutes ces données est presque impossible. Au cours de la dernière décennie, la nécessité de traiter ces ensembles de données de plus en plus volumineux et non structurés a conduit l'analyse traditionnelle des données à se transformer en "Big Data" [5].

4. Structure de Big Data

Une structure de données concerne l'ensemble des valeurs de données, de relations entre elles et de fonctions ou d'opérations qui peuvent être effectuées sur les données. Il s'agit d'une technique de stockage et d'organisation des données dans un environnement informatique qui permet d'y accéder et de les modifier facilement [7].

Dans le contexte du Big Data, les données récoltées, stockées et traitées peuvent être originaires de plusieurs domaines et être formées par plusieurs sources de données hétérogènes, ce qui donne lieu à une masse de données de différentes sortes (figure 2) : (1) structurées, (2) non structurées, (3) semi-structurées et (4) quasi-structurées [7].

1. Données structurées

Les données structurées sont des données qui ont un format et une longueur prédéterminés, qui sont simples à stocker et à analyser et qui sont bien organisées. Cela implique que les données ont été structurées en une forme reconnaissable qui leur permet de répondre à des requêtes pour récupérer des informations à des buts organisés [8].

Une base de données relationnelle, telle que SQL (Structured Query Language), est un excellent exemple de données structurées, elle contient par exemple des dates, des nombres organisés, des groupes de mots et des nombres [8].

2. Données non structurées

Les données non structurées sont des informations sous divers formats qui ne correspondent pas aux modèles de données typiques et qui ne peuvent pas être utilisées dans une base de données relationnelle traditionnelle. Cela rend le traitement et l'analyse des données non structurées extrêmement compliqués et longs [8].

Les données non structurées n'ont pas de structure précise. Des images, des textes, des courriels et d'autres types de données qui ne sont pas stockés dans une base de données sont des exemples de données non structurées [9].

3. Données semi structurées

Les données semi-structurées sont des données irrégulières qui peuvent être incomplètes et dont la structure change rapidement ou de manière imprévisible, mais qui ne se conforment pas à un format fixe ou explicite [10].

Selon Hanig, Schierle et Trabold, le type de données semi-structurées permet de regrouper en un ensemble unique des informations provenant de nombreuses sources, chacune ayant des qualités connexes mais distinctes, comme les courriels, le XML et les fichiers Doc [10].

4. Données quasi structurées

Des données textuelles avec des formats aléatoires qui peuvent être formatées à l'aide d'outils spécifiques [11], par exemple, les journaux du serveur web, Internet Explorer, un fichier journal qui est automatiquement créé et maintenu par un serveur et qui contient une liste d'activités [12].

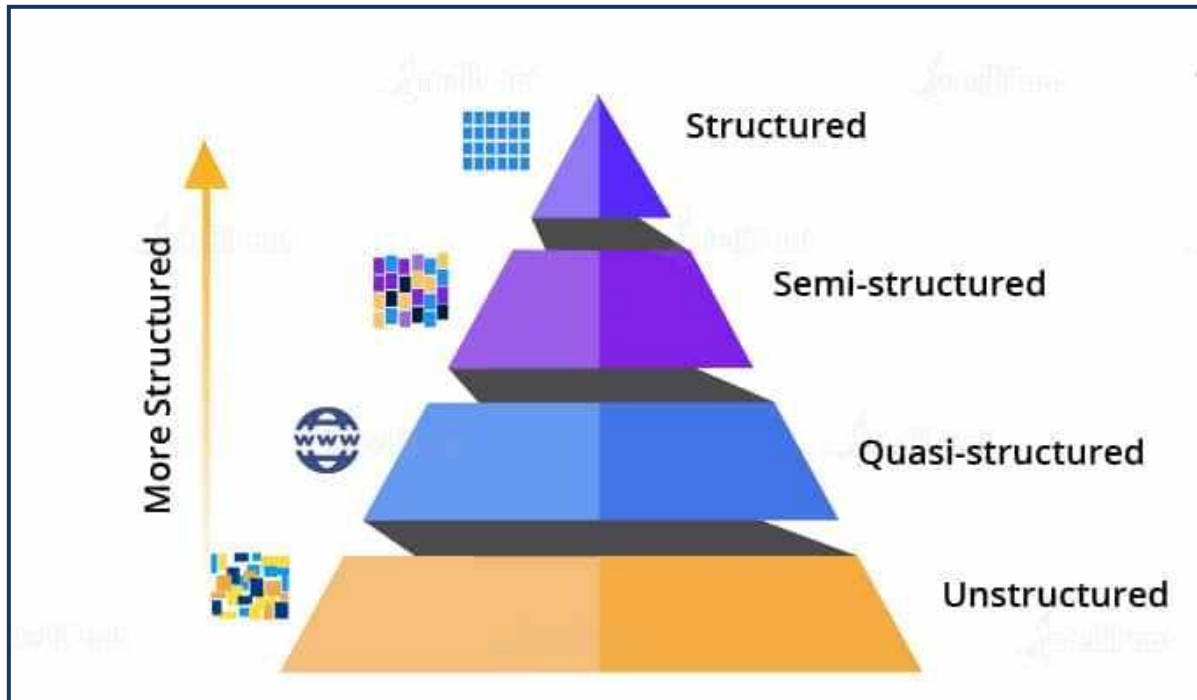


Figure 2 : Structure des données massives [12].

5. Caractéristique Big Data

Les caractéristiques des Big Data ont connu plusieurs évolutions et changements, dont les principaux sont les suivants (figure 3) :

1. Volume

Le volume est considéré comme le premier critère du Big Data, car il correspond à la masse des données produite à travers nous accès en générale, ce dernier se développe et augmente à une cadence exponentielle [11].Donc on peut conclure qu'il fait référence à la quantité de données produites chaque unité du temps par les utilisateurs (compagnies, personnes, etc.).

2. Vitesse

Ce critère de vitesse fait référence à la capacité de traiter des ensembles de données en un temps record. Ceci permet de créer un service directement basé sur les interactions actuelles. Il est donc lié à la vitesse ou à la fréquence de génération des données, ainsi qu'à la vitesse de partage et de circulation des données [13].

3. Variété

La diversité est un critère important dans la définition des données massives, elle représente un véritable défi technologique. Au fait, les Big Data proviennent d'une variété de sources et de formats (vidéos, sons, documents, journaux, bases de données, données de capteurs, itinéraires, etc.). Les ensembles des Big Data se composent de données structurées et non structurées, publiques ou privées, locales ou distantes, partagées ou confidentielles, complètes ou incomplètes [14].

4. Véracité

La véracité fait appel à la précision et à la fiabilité des données. Pour retirer la valeur des données, celles-ci doivent être nettoyées afin d'éliminer le bruit. Ce nettoyage des données est important afin de filtrer les données incorrectes et erronées [15].Comme le Big Data permet de collecter un nombre indéfini et plusieurs formes de données, il est difficile de justifier l'authenticité des contenus, elle apporte un peu d'ordre a tous cela en associant les analytiques qui correspondent aux besoins des utilisateurs [16].

5. Valeur

La valeur des données fait référence à l'utilité des données en fonction de l'objectif visé. L'objectif ultime de l'ensemble du système d'analyse des méga données est d'extraire cette valeur des données [13], elle est également liée à la véracité ou à l'exactitude des données. Pour certaines parties, la valeur dépend également de la rapidité avec laquelle nous pouvons traiter les données. Alors, La valeur est le profit qu'on puisse gagner de l'utilisation des données massives [13].

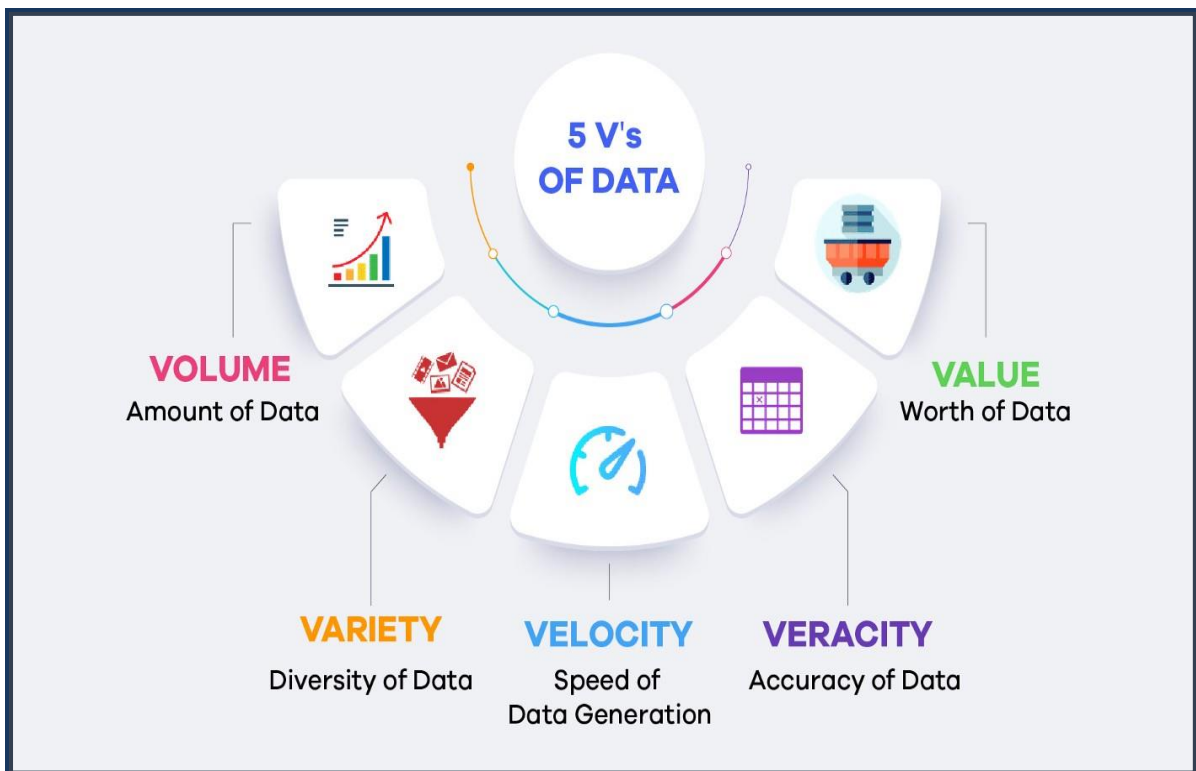


Figure 3: Les 5V des Big Data [17].

6. Clés des Big Data

Selon les recherches effectuées par le professeur " Gerard Gerry" de l'université de Georgetown et d'autres chercheurs dans le domaine du Big Data, il existe cinq sources principales de Big Data [63] :

1. Les données publiques

Les données publiques font référence aux données détenues par les administrations, les organismes publics et les collectivités locales qui peuvent potentiellement être utilisées dans des projets commerciaux ou des applications de gestion à grande échelle [63].

2. Les données privées

Les données privées sont détenues par des entreprises privées et par des particuliers et reflètent des informations personnelles qui ne peuvent être facilement obtenues auprès de sources publiques [63].

3. Les échappements des données

Les échappements ou les fuites de données sont des données ambiantes non essentielles qui sont collectées passivement et qui n'ont que peu ou pas de valeur pour le partenaire de collecte de données d'origine. Ces données peuvent être recombinaisonnées avec des informations provenant d'autres sources pour produire une nouvelle source de valeur [63].

4. Les données communautaires

Les données communautaires (communication sociale) sont la distillation de données non structurées en réseaux dynamiques qui capturent les tendances sociales [63].

5. Les données d'auto-quantification

Les données auto-quantifiées sont des données révélées par les individus en quantifiant leurs activités et leurs comportements personnels [63].

7. Processus de traitement des Big Data

Parmi les étapes du traitement du Big Data, nous mentionnons les suivantes :

1. Génération de données

La génération de données est la première étape du Big Data. Étant donné les données Internet comme exemple, une quantité énorme de données en termes de recherche d'entrées, de messages de forums Internet, et d'enregistrements de discussion, est générée. Ces données sont étroitement liées à la vie quotidienne des gens et présentent des caractéristiques similaires de grande valeur et de faible densité. Ces données Internet peuvent être sans valeur individuellement, mais, grâce à l'exploitation de méga données accumulées, des informations utiles telles que les habitudes et les loisirs des utilisateurs peuvent être identifiées, et il est même possible de prévoir les comportements et les humeurs émotionnelles des utilisateurs [18].

Les Big Data peuvent être générées par des sources de données longitudinales et/ou distribuées telles que des capteurs, des vidéos, des flux de clics et d'autres sources de données. Ces informations dépassent de loin les capacités des infrastructures informatiques des entreprises existantes [18].

2. Acquisition

L'acquisition des Big Data est la deuxième phase du système Big Data, et elle comprend la collecte, la transmission et le prétraitement des données. Après avoir collecté les données brutes lors de l'acquisition des Big Data, nous utiliserons un mécanisme de transmission efficace pour les envoyer vers un système de gestion de stockage approprié afin de prendre en charge différentes applications analytiques [18].

Les ensembles de données collectées peuvent parfois inclure beaucoup de données redondantes ou inutiles, ce qui augmente inutilement l'espace de stockage et affecte l'analyse ultérieure des données. Par exemple, une redondance élevée est très courante parmi les ensembles de données collectées par des capteurs pour la surveillance de l'environnement. La technologie de compression des données peut être appliquée pour réduire la redondance. Par conséquent, les opérations de prétraitement des données sont indispensables pour assurer un stockage et une exploitation efficaces des données [18].

3. Prétraitement des données

En raison de la grande variété des sources de données, les ensembles de données collectées varient en termes de bruit, de redondance, de cohérence, etc. le stockage de données sans intérêt constitue sans aucun doute un gaspillage [18].

En outre, certaines méthodes d'analyse ont de sérieuses exigences en matière de qualité des données. Par conséquent, afin de permettre une analyse efficace des données, nous devons prétraiter les données dans de nombreuses circonstances pour intégrer les données provenant de différentes sources, ce qui peut non seulement réduire les frais de stockage, mais aussi améliorer la précision de l'analyse. Certaines techniques de prétraitement des données relationnelles sont présentées ci-dessous [18].

A) L'intégration

L'intégration des données est la base de l'informatique commerciale moderne. Elle consiste à combiner des données provenant de diverses sources et à fournir aux utilisateurs

une vue cohérente des données [19]. Il s'agit d'un domaine de recherche établi pour les bases de données traditionnelles [19].

B) Le nettoyage

Le nettoyage des données est un processus visant à identifier les données incorrectes, incomplètes ou non raisonnables, puis à modifier ou supprimer ces données afin d'améliorer la qualité [20].

En général, le nettoyage des données comprend cinq procédures complémentaires : définir et déterminer les types d'erreurs, rechercher et identifier les erreurs, corriger les erreurs, documenter les exemples d'erreurs et les types d'erreurs, et modifier les procédures de saisie des données pour réduire les erreurs futures [20].

Pendant le nettoyage, les formats de données, la complétude, la rationalité et la restriction sont inspectés. Le nettoyage des données est d'une importance essentielle pour maintenir la cohérence des données, qui est largement appliquée dans de nombreux domaines, tels que la banque, l'assurance, le commerce de détail, les télécommunications et le contrôle du trafic [20].

C) Élimination de la redondance

La redondance des données fait référence aux répétitions de données, qui sont courantes dans les grands ensembles de données. La redondance des données peut augmenter les coûts de transmission inutiles des données et provoquer des défauts du système de stockage, comme le gaspillage de l'espace de stockage, ce qui entraîne une incohérence des données, une réduction de la fiabilité des données et des dommages aux données. En conséquence, diverses méthodes de réduction de la redondance, telles que la détection de la redondance, le filtrage des données et la compression des données, ont été proposées [18].

Ces méthodes peuvent être applicables à divers ensembles de données ou environnements d'application. Cependant, la réduction des redondances peut avoir des conséquences négatives. La compression et la décompression des données, par exemple, augmentent la charge de calcul. Par conséquent, les coûts et les avantages de la réduction des redondances doivent être soigneusement équilibrés [18].

4. Stockage

La croissance explosive des données entraîne des exigences plus strictes en matière de stockage et de gestion. Dans cette section, nous nous concentrons sur le stockage des données volumineuses [18].

Le stockage des Big Data fait référence au stockage et à la gestion d'ensembles de données à grande échelle tout en assurant la fiabilité et la disponibilité de l'accès aux données. Nous passerons en revue des problèmes importants, notamment les systèmes de stockage massif, les systèmes de stockage distribué et les mécanismes de stockage des données volumineuses [18].

D'une part, l'infrastructure de stockage doit fournir un service de stockage d'informations avec un espace de stockage fiable ; d'autre part, elle doit fournir une interface d'accès puissante pour l'interrogation et l'analyse d'une grande quantité de données [18].

Traditionnellement, en tant qu'équipement auxiliaire du serveur, le dispositif de stockage de données est utilisé pour stocker, gérer, rechercher et analyser données avec des SGBDR structurés. Avec la forte croissance de données, le dispositif de stockage de données devient de plus en plus important, et de nombreuses sociétés d'Internet recherchent une grande capacité de stockage pour être compétitives. Par conséquent, il existe un besoin impérieux de recherche sur le stockage des données [18].

A) Système de stockage

Divers systèmes de stockage émergent pour répondre aux exigences des données massives. Les technologies de stockage massives existantes peuvent être classées comme stockage en attachement direct (DAS : Direct Attached Storage) et stockage réseau, tandis que le stockage réseau peut être classé dans le stockage en réseau (NAS : Network Attached Storage) et la zone de stockage Réseau (SAN : Storage Area Network) [18].

Dans le système DAS, des différents disques durs sont directement connectés aux serveurs, la gestion des données est centrée sur le serveur, de sorte que les dispositifs de stockage sont des équipements périphériques dont chacun prend une certaine quantité de ressources d'E/S et est géré par une application individuelle [18].

Pour cette raison, il ne convient que pour interconnecter des serveurs de petite taille. Toutefois, en raison de sa faible évolutivité, le DAS sera efficace lorsque la capacité de

stockage sera augmentée, c'est-à-dire lorsque l'évolutivité et l'extensibilité seront très limitées. Ainsi, DAS est principalement utilisé dans les ordinateurs personnels et les petites les serveurs [18].

Dans le système **NAS**, Le stockage en réseau consiste à utiliser le réseau pour fournir aux utilisateurs une interface d'union pour l'accès et le partage des données. L'équipement de stockage réseau comprend des équipements d'échange de données spéciaux, une baie de disques, une bibliothèque et d'autres supports de stockage, ainsi que des logiciels de stockage spéciaux. Il se caractérise par une forte extensibilité [18].

Le SAN est spécialement conçu pour le stockage de données avec un réseau évolutif et à large bande passante, par exemple un réseau à haut débit avec des connexions en fibre optique. Dans un réseau SAN, la gestion du stockage des données est relativement indépendante à l'intérieur d'un réseau local de stockage, où la commutation des données par trajets multiples entre tous les nœuds internes est utilisée pour obtenir un degré maximal de partage et de gestion des données [18].

B) Système de stockage distribué

Le premier défi posé par le Big Data est de savoir comment développer un système de stockage distribué à grande échelle pour un traitement et une analyse efficace des données. Pour utiliser un système distribué pour le stockage des données massives, les facteurs suivants doivent être pris en considération la cohérence, la disponibilité, tolérance de partition [21].

Eric Brewer a proposé une théorie CAP [21, 22] en 2000, qui indiquait qu'un système distribué ne pouvait pas répondre simultanément aux exigences de cohérence, de disponibilité, et tolérance de partition ; au plus deux des trois exigences peuvent être satisfaites simultanément. Seth Gilbert et Nancy Lynch du MIT a prouvé l'exactitude de la théorie CAP en 2002. Depuis la cohérence, la disponibilité et la partition la tolérance n'a pas pu être atteinte simultanément, nous pouvons avoir un système CA en ignorant la tolérance de partition, un système CP en ignorant la disponibilité, et un système AP qui ignore la cohérence, selon différents objectifs de conception. Les trois systèmes sont discutés dans ce qui suit.

C) Mécanisme de stockage

De nombreuses recherches sur les méga données favorisent le développement des mécanismes de stockage pour les méga données. Les mécanismes de stockage existants

peuvent être classés en trois niveaux : systèmes de fichiers, bases de données, et modèles de programmation.

Les systèmes de fichiers sont à la base des applications de niveaux supérieurs. Le GFS (Google File Système) de Google est une extension du système de fichiers distribuée pour prendre en charge les données à grande échelle, qui utilise des serveurs qui marchent très bien pour atteindre la tolérance aux pannes et fournir aux clients des services de haute performance [23].

En outre, HDFS et Kosmosfs sont des dérivés des codes open source de GFS. Microsoft a développé Cosmos [23] pour soutenir ses activités de recherche et de publicité. Facebook utilise Haystack [24] pour stocker la grande quantité de photos de petite taille.

La base de données NoSQL : Les bases de données NoSQL se caractérisent par des modes flexibles, la prise en charge d'une copie simple et facile, une API simple, une cohérence éventuelle et la prise en charge de gros volumes de données. Les bases de données NoSQL sont en train de devenir la technologie de base pour le Big Data [25].

Les trois principales bases de données NoSQL sont : Les bases de données à valeur clé, les bases de données orientées colonnes et les bases de données orientées documents, chacune étant basée sur un certain modèle de données [26,27].

Les modèles de programmation : des recherches ont également porté sur les modes de programmation pour des tâches de calcul plus complexes, par exemple, les calculs itératifs, les calculs en mémoire tolérants aux pannes, les calculs incrémentaux et le contrôle de flux [18].

5. Analyse des données Big Data

L'analyse de données traditionnelle signifie l'utilisation des méthodes statistiques appropriées pour analyser des données massives, concentrer, extraire et affiner des données utiles cachées dans un lot d'ensembles de données chaotiques et identifier la loi inhérente au sujet, afin de maximiser la valeur des données [18].

L'analyse des données joue un rôle d'orientation considérable dans l'élaboration des plans de développement d'un pays, la compréhension des demandes des clients en matière de commerce et la prévision des tendances du marché pour les entreprises. L'analyse des méga

données peut être considérée comme la technique d'analyse d'un type particulier de données [18].

Par conséquent, de nombreuses méthodes d'analyse de données traditionnelles peuvent encore être utilisées pour l'analyse de méga données. Plusieurs méthodes d'analyse de données traditionnelles représentatives sont examinées ci-après, dont beaucoup proviennent de la statistique et de l'informatique.

8. Défis du Big Data

Dans le monde des données massives, nous trouvons de nombreuses difficultés, dont nous citons les suivantes :

1. Représentation des données :

De nombreux ensembles de données varient en termes de forme, de structure, de sémantique, d'organisation, de granularité et d'accessibilité. L'objectif de la représentation des données est de les rendre plus intelligibles pour l'analyse informatique et l'interprétation par l'utilisateur. Néanmoins, une représentation incorrecte des données réduit la valeur des données d'origine et peut même entraver la réussite de leur analyse [28].

Une représentation efficace des données doit tenir compte de la structure, de la classe et du type des données, ainsi que des technologies intégrées, afin de permettre des opérations efficaces sur divers ensembles de données [28].

2. Réduction de la redondance et compression des données

En général, il existe un niveau élevé de redondance dans les ensembles de données, La réduction de la redondance et la compression des données permettent de réduire le coût indirect de l'ensemble du système en partant du principe que les valeurs potentielles des données ne sont pas affectées. Par exemple, la plupart des données générées par les réseaux de capteurs sont fortement redondantes, et peuvent être filtrées et compressées à des ordres de grandeur [28].

3. Mécanisme analytique

Le système analytique du Big Data doit traiter des masses de données hétérogènes dans un temps limité. Cependant, les SGBDR traditionnels sont strictement conçus avec un manque d'évolutivité et d'extensibilité, ce qui ne pouvait pas répondre aux exigences de performances. Les bases de données non relationnelles ont montré leurs avantages uniques dans le traitement des données non structurées et ont commencé à se généraliser dans l'analyse des méga données. Même ainsi, il existe encore des problèmes de bases de données non relationnelles dans leurs performances et leurs applications particulières. Nous trouverons une solution de compromis entre les SGBDR et les bases de données non relationnelles [28].

4. Confidentialité des données

La plupart des fournisseurs ou des actionnaires de services Big Data ne peuvent actuellement pas gérer et analyser efficacement des ensembles de données aussi volumineux en raison de leurs capacités limitées. Ils doivent s'en remettre à des professionnels ou à des outils pour analyser ces données, ce qui augmente les risques potentiels pour la sécurité [28].

5. Gestion de l'énergie

La consommation d'énergie des systèmes informatiques centraux a attiré beaucoup d'attention tant du point de vue de l'économie que de l'environnement. Avec l'augmentation du volume de données et des demandes analytiques, le traitement, le stockage et la transmission des méga données consommeront inévitablement de plus en plus d'énergie électrique [28].

Par conséquent, un mécanisme de contrôle et de gestion de la consommation d'énergie au niveau du système doit être établi pour les méga données tout en garantissant l'évolutivité et l'accessibilité [28].

6. Extensibilité et évolutivité

Le système d'analyse des données doit être capable de traiter les jeux de données actuels et futurs. L'algorithme analytique doit être capable de traiter des jeux de données de plus en plus volumineux et complexes [28].

7. Coopération

L'analyse des méga données est une recherche interdisciplinaire, qui exige que des experts de différents domaines coopèrent pour exploiter le potentiel des méga données.

Une architecture complète de réseau de méga données doit être établie pour aider les scientifiques et les ingénieurs de divers domaines à accéder à différents types de données et à utiliser pleinement leur expertise, afin de coopérer pour atteindre les objectifs analytiques [28].

9. Relation entre les méga données et l'IoT

Dans le paradigme de l'IoT, une énorme quantité de capteurs sont intégrés dans divers appareils et machines dans le monde réel. Ces capteurs déployés dans différents domaines peuvent recueillir divers types de données, telles que des données environnementales, géographiques, astronomiques et logistiques. Les équipements mobiles, les installations de transport, les installations publiques et les appareils ménagers peuvent tous être des équipements d'acquisition de données dans l'IoT, comme le montre la figure 4 [18].

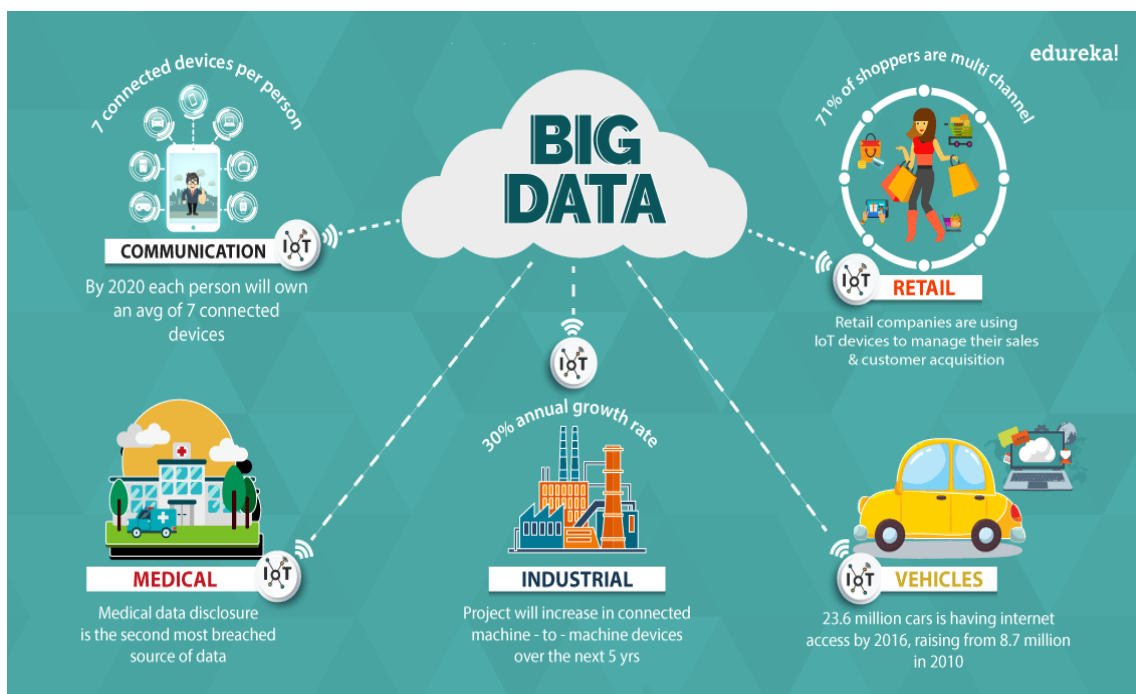


Figure 4 : Relation entre Big Data et internet des objets [29].

Le Big Data généré par l'IoT a des caractéristiques différentes par rapport au Big Data général en raison des différents types de données collectées, dont les caractéristiques les plus classiques sont l'hétérogénéité, la variété, la fonctionnalité non structurée, le bruit et la forte redondance. Bien que les données IoT actuelles ne soient pas la partie dominante des méga données, d'ici 2030, la quantité de capteurs atteindra un billion, puis les données IoT seront la partie la plus importante des méga données, selon les prévisions [18].

À l'heure actuelle, la capacité de traitement des données de l'IoT a pris du retard par rapport aux données collectées et il est extrêmement urgent d'accélérer l'introduction des technologies du Big Data pour favoriser le développement de l'IoT. De nombreux opérateurs de l'IoT réalisent l'importance du Big Data puisque le succès de l'IoT repose sur l'intégration efficace du Big Data et du cloud computing. Le déploiement généralisé de l'IoT fera également entrer de nombreuses villes dans l'ère du Big Data [18].

Il existe un besoin impérieux d'adopter le Big Data pour les applications IoT, alors que le développement du Big Data est déjà à la traîne. Il a été largement reconnu que ces deux technologies sont interdépendantes et doivent être développées conjointement : d'une part, le déploiement généralisé de l'IoT entraîne la forte croissance des données à la fois en quantité et en catégorie, offrant ainsi la possibilité d'appliquer et de développer le Big Data, d'autre part, l'application de la technologie du Big Data à l'IoT accélère également les avancées de la recherche et les modèles commerciaux de l'IoT [18].

10. Conclusion

Le volume de Big Data est déjà massif, mais il devrait augmenter de façon exponentielle avec l'arrivée de nouvelles technologies, il faut donc travailler à trouver des solutions simples et pratiques pour stocker les données volumineuses et en assurer l'accès pour les traiter et les analyser.

Les données massives peuvent faire progresser la recherche scientifique dans tous les domaines pour améliorer et faciliter la vie humaine.

Le « Big Data » aide les architectes, les chercheurs, les ingénieurs et les écologistes à construire des villes intelligentes, qui facilitent la vie quotidienne des gens tout en préservant l'environnement, où la santé, le gouvernement, l'éducation et l'économie utilisent de l'intelligence. Dans le chapitre 2, nous aborderons les notions de ville intelligente.

CHAPITRE II : SMART CITY

1. Introduction

Avec l'énorme développement technologique que nous avons réalisé ces dernières années et la grande, plusieurs nouveaux concepts ont vu le jour aujourd'hui. Parmi les quels, on peut citer le concept de Smart city ou villes intelligentes en français. Elles consistent principalement en l'optimisation des coûts, de l'organisation, et de la commodité des habitants. Les villes intelligentes ont pour objectif aussi de répondre aux exigences des utilisateurs par sa disponibilité, sa compatibilité environnementale, sa rentabilité et sa compatibilité sociale.

Dans ce chapitre, nous présentons quelques généralités des villes intelligentes, leurs objectifs, leurs outils et leurs domaines d'application dans la vie quotidienne.

2. Définitions

Littéralement, le terme Smart city désigne les « villes intelligentes ». Il s'agit d'un terme anglais utilisé pour définir les villes qui utilisent les technologies de l'information et de la communication pour améliorer la qualité de divers services urbains. D'autres termes similaires ont été utilisés à savoir : ville connectée, cyber ville, ville numérique [30].

Plusieurs définitions ont été données à la notion de Smart city, Parmi les, nous citons les définitions suivantes :

- Les initiatives Smart Cities tentent d'améliorer les performances urbaines en utilisant les données, l'information et les technologies de l'information, pour fournir des services plus efficaces aux citoyens, surveiller et optimiser les infrastructures existantes, accroître la collaboration entre les différents acteurs économiques et encourager des modèles commerciaux innovants dans le secteur privé et les secteurs publics [31].
- Les villes intelligentes sont définies de plusieurs façons. En substituant des adjectifs alternatifs à « intelligent », tels que « intelligent » ou « numérique », une variété de variantes conceptuelles peut être produite. Le terme « ville intelligente » est un terme nébuleux qui s'applique de diverses manières. Il n'y a pas de définition universelle d'une ville intelligente [32].

- Selon Harrison et al. (2010) dans un document d'entreprise d'IBM, le terme « ville intelligente » fait référence à une « ville instrumentée, interconnectée et intelligente ». Le terme « instrumenté » fait référence à la capacité de capturer et d'intégrer des données réelles en direct via des capteurs, des compteurs, des appareils, des appareils personnels et d'autres capteurs similaires. « Interconnecté » fait référence à l'incorporation de ces données dans une plate-forme informatique permettant la communication de ces informations entre les différents services de la Ville. L'utilisation de services complexes d'analyse, de modélisation, d'optimisation et de visualisation pour prendre de meilleures décisions opérationnelles est qualifiée d'« intelligente » [33].

3. Historique

Le terme Smart city est apparu pour la première fois en 1990. L'intérêt s'est alors porté sur l'avancement des technologies de l'information et de la communication en relation avec les infrastructures urbaines modernes [30].

Selon Breux et Diaz (2017), le mot « Smart city » a trois racines :

- En 2009, IBM a popularisé cette expression après avoir vu les villes comme un marché potentiel viable pour le développement des technologies de l'information et de la communication [34].
- Nam et Pardo (2011) ont découvert une multitude d'expressions parallèles à la ville intelligente et les ont divisées en trois catégories : humaine, institutionnelle et technologique. Selon ces auteurs, une ville est dite intelligente lorsque les investissements favorisent une croissance durable et améliorent la qualité de vie par une gouvernance participative, en fonction de l'interaction entre ces trois caractéristiques [35].
- Selon Sajhau (2017), le sujet des villes intelligentes prend une nouvelle importance lorsque trois tendances actuelles entrent en collision : l'urbanisation croissante, les défis environnementaux et la révolution numérique [36].

D'après les recherches, la notion de ville intelligente a évolué au fil du temps, les définitions devenant plus précises, puis plus axées sur l'humain, mais se concentrant presque toujours sur les mêmes thèmes fondamentaux : la ville, la technologie et la durabilité.



Figure 5 : Smart city [37].

4. Les objectifs de la Smart city

Les objectifs de la Smart city sont étroitement liés au développement durable qui est une croissance qui satisfait les demandes des générations actuelles sans compromettre la capacité des générations futures à satisfaire leurs propres besoins. Cette idée est fondée sur la notion de « besoins », notamment les besoins essentiels des plus pauvres. C'est aussi une analyse des limites que nos techniques et notre organisation sociale actuelles imposent à l'environnement [38].

Trois objectifs importants sont définis :

- Au-delà de la satisfaction des besoins humains fondamentaux, le développement durable tente de répondre au désir d'améliorer la qualité de vie des personnes et, dans le cas des villes intelligentes, la qualité de vie des habitants [38].
- Pour atteindre les exigences fondamentales, une croissance économique est nécessaire. Cette expansion économique doit se faire dans le respect des lois de la durabilité, et pas seulement pour répondre aux exigences [38].

- L'environnement doit être respecté pour que la croissance soit durable, il s'agit d'éviter la surexploitation des ressources. « Par ailleurs, si les progrès technologiques peuvent atténuer certains problèmes, ils peuvent aussi en générer de nouveaux [38]. Par conséquent, ce sujet est particulièrement délicat dans le cas des villes intelligentes, qui tirent leur essence du développement technologique exponentiel tout en poursuivant la durabilité comme objectif.

5. Les outils de la Smart city

La Smart city utilise des technologies modernes pour créer des solutions innovantes en collaboration avec tous les acteurs urbains.

1. La technologie

La technologie c'est l'ensemble des théories, des procédés et des méthodes grâce auxquels il est possible de profiter de façon pratique de la connaissance scientifique. Le mot désigne alors les connaissances permettant de fabriquer des objets et de modifier l'environnement, dans le but de satisfaire les besoins humains [39].

Aussi, le terme technologie fait référence à un ensemble d'instruments qui permettent la formation de villes intelligentes, tels que les réseaux intelligents, les sources d'énergie renouvelables, les nouveaux types de carburants, les nouveaux matériaux de construction, etc... [40].

La ville intelligente utilise les « nouvelles » technologies comme des facilitateurs pour rendre les villes plus pratiques, plus écologiques et plus agréables à vivre. Les villes deviennent plus intelligentes au fur et à mesure de l'émergence de nouvelles technologies, intégrant des aspects de la vie urbaine tels que l'économie, l'éducation, la démocratie, les infrastructures, les transports, l'environnement, la sécurité et la qualité de vie [40].

2. Numérique

Les technologies de l'information et de la communication, en particulier, constituent un sous-ensemble des technologies numériques. Elles permettent la création de services qui aident les utilisateurs à analyser et à anticiper leurs demandes. Dans un premier temps, elles permettent d'organiser les villes autour de systèmes intelligents qui collectent, analysent et

traitent les données générées par les différents acteurs urbains (tels que les systèmes et services municipaux ou les citoyens, etc.) [41].

Ensuite, à partir de ces données, des modèles d'analyse peuvent être créés afin de détecter à l'avance les problèmes potentiels et d'en atténuer les conséquences. De plus, ils s'assurent que toutes ces données sont connectées, ce qui permet une prise de décision collaborative en temps réel [41].

Enfin, des outils et des logiciels doivent être développés afin de collecter, analyser, traiter et intégrer ces données.

3. Internet of Things

L'internet des objets (IOT) est un réseau d'objets physiques, selon la définition la plus largement acceptée. L'internet s'est transformé en un réseau d'appareils de tous types et de toutes tailles, y compris des véhicules, des téléphones intelligents, des appareils ménagers, des jouets, des caméras, des instruments médicaux et des systèmes industriels, des animaux, des personnes et des bâtiments, tous connectés, communiquant et partageant des informations sur la base de protocoles prédéterminés afin de réaliser des réorganisations intelligentes, le positionnement, le traçage, la sécurité et le contrôle, et même la surveillance en ligne personnelle en temps réel, la mise à niveau et la propulsion [42].

Ces objets interconnectés sont utilisés pour recueillir, stocker et envoyer des données sur notre environnement en temps réel. Ces données sont examinées et utilisées pour améliorer la gestion publique de la ville dans le cadre de la ville intelligente [42].

4. Open Data

Il apparaît nécessaire de concevoir un système de gestion des données en même temps que l'IOT afin de recueillir, stocker, analyser et diffuser les informations dans le cadre d'une politique de gestion Open Data préétablie [43].

C'est l'aspect le plus important des Open Data. Ce mot désigne toute donnée dont l'accès est entièrement libre et sans restriction, ainsi que son exploitation et sa réutilisation [43].

Open Data doivent répondre à trois critères : disponibilité, réutilisation et distribution, ainsi que participation universelle [43]. L'interopérabilité, ou la capacité à combiner des sources de données disparates, est définie par ces trois caractéristiques.

6. Les acteurs de la Smart city

Parmi les nombreuses réflexions à mener sur les villes intelligentes, il convient de se demander si une redéfinition du rôle et du sens des villes est nécessaire pour justifier les hyper technologies. En d'autres termes, la ville intelligente est une collaboration intersectorielle entre une variété de parties prenantes (entreprises privées, agences gouvernementales, scientifiques, citoyens, etc.) ayant des objectifs divers [44].

Pour identifier les nombreux participants importants de la ville intelligente, nous utiliserons le modèle de la « quadruple hélice » [45]. Selon ces auteurs, quatre types d'acteurs sont identifiés dans un projet de la ville intelligente : (1) le public (autorités et administration locales), (2) le secteur des affaires (entreprises locales et étrangères), (3) les centres de recherche et universitaires, et (4) les résidents. Ces quatre groupes d'acteurs collaborent dans un cadre urbain dynamique où l'engagement des citoyens sert de ciment entre l'université, l'industrie et le gouvernement. L'interaction entre ces individus définit le succès d'une ville en tant que ville intelligente [45].

7. Les domaines de la Smart city

Le professeur Rudolph Giffinger, de l'Université technologique de Vienne en Autriche, a rédigé une étude en 2007 identifiant les six dimensions d'une ville intelligente [46]. Ce modèle, qui est largement utilisé et adopté comme référence dans le monde entier (tant par les universitaires que par les praticiens), propose une classification des champs d'application de la ville intelligente en six domaines : Économie intelligente, Mobilité intelligente, Gouvernance intelligente, Environnement intelligent, Personnes intelligentes et Vie intelligente (figure 6).

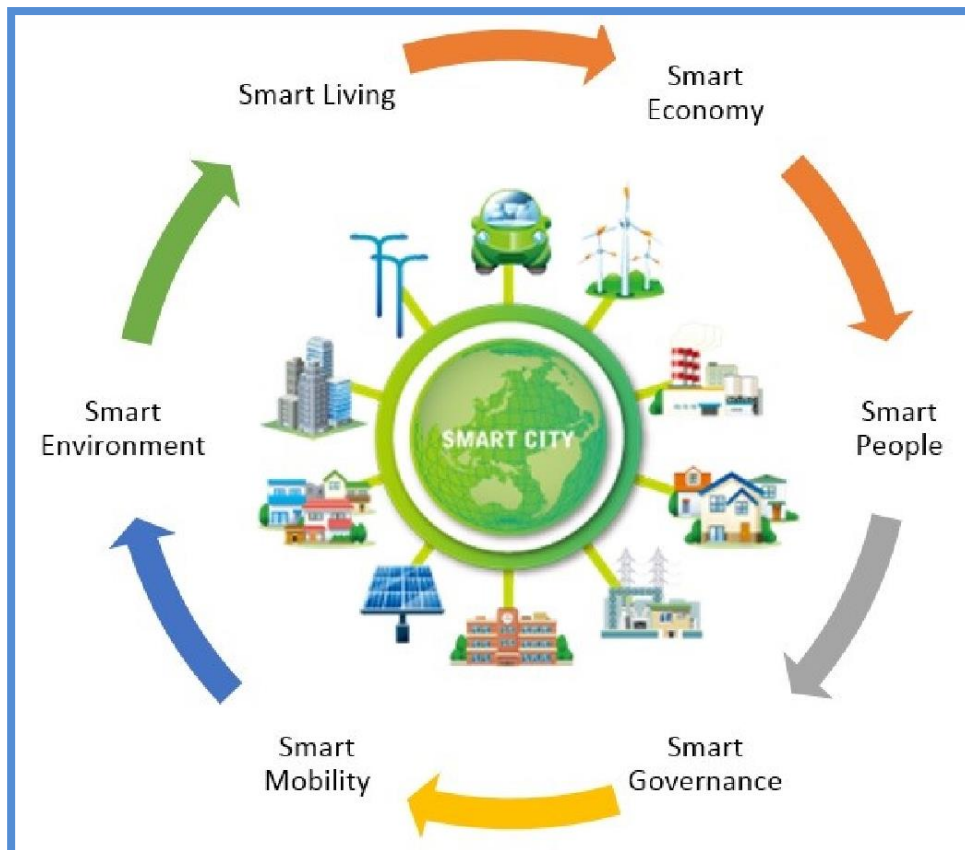


Figure 6 : Les Domaines de la Smart city [55].

1. Smart Economy

Le terme « Smart economy » ou « économie intelligente » a une définition large car il fait référence à la dimension économique d'une ville intelligente. Avec l'intégration des technologies du futur dans l'industrie, l'essor des plateformes numériques et la révolution industrielle induite par l'Internet des objets, de nouvelles exigences et attentes sont naturellement apparues dans les mentalités des consommateurs [47].

La Smart economy est une économie plus inventive et intelligente dans laquelle les entreprises doivent « faire plus avec moins » tout en répondant à l'évolution des attentes et des habitudes des consommateurs. Plus précisément, l'économie intelligente implique la fourniture de nouveaux services, la fabrication de nouveaux produits, le développement de nouveaux modèles économiques, l'interaction du monde local et du monde globalisé, et l'établissement de nouveaux modèles d'affaires [47].

L'objectif principal de l'économie intelligente est de stimuler l'économie de la ville afin de répondre aux attentes des citoyens tout en la rendant plus compétitive et plus attrayante pour les consommateurs [47].

2. Smart Mobility

La mobilité intelligente est un sous-ensemble de la technologie des villes intelligentes qui vise à fournir des solutions pratiques et à long terme aux problèmes de mobilité urbaine. Nous parlons de mobilité intelligente parce qu'elle repose sur l'application de nouvelles technologies de télécommunications, d'économie collaborative et de bases de données en ligne pour améliorer les infrastructures et les modes de transport, permettant une mobilité plus souple, plus fiable et plus multimodale [48].

L'objectif majeur de la mobilité intelligente est de rendre la vie des habitants plus facile, plus rapide et plus confortable tout en limitant l'utilisation des ressources et en protégeant l'environnement et couvre divers sous objectifs (la sécurité routière, La diminution d'embouteillage routière en ville, L'amélioration du stationnement, etc...) [48].

3. Smart Environment

Le mot de Smart environment est présente par les professeurs : Leonidas Anthopoulos, Marijn Janssen et Vishanth Weerakkody on 2017 [49], comme la protection et la gestion des ressources naturelles par l'innovation et l'intégration technologiques (systèmes de gestion de la consommation d'énergie, contrôle des émissions, recyclage, capteurs de surveillance de la pollution, etc.)

L'objectif de l'environnement intelligent comprend de nombreuses filières, notamment (Smart environment, Smart energy, Smart buildings, Smart water et Smart waste).

- **La gestion intelligente de l'énergie** : vise une production d'énergie plus respectueuse de l'environnement, une consommation d'énergie plus faible, un modèle de consommation d'énergie plus prévisible et un réseau de distribution plus résilient [50].

- **Les bâtiments intelligents** utilisent un grand nombre de capteurs pour générer des données en temps réel (comme la température, l'humidité et la lumière) qui offrent des informations sur l'occupation et les conditions de vie. Ces données sont utilisées pour suivre les systèmes de gestion du bâtiment comme le refroidissement, la ventilation et l'éclairage, et pour ajuster la quantité d'énergie utilisée en fonction du nombre de personnes dans le bâtiment ou d'autres facteurs [51].

- **Les systèmes intelligents de gestion de l'eau et des déchets** visent à améliorer la qualité de l'eau, réduisant les déchets et à diminuer le nombre de camions poubelles et leurs trajets. Divers systèmes intelligents ont déjà vu, qui autorise, par exemple, l'utilisation de capteurs pour vider les conteneurs lorsqu'ils sont pleins.

4. Smart People

Smart people est considéré comme une des trois dimensions clés de la Smart city [35], les villes intelligentes sont centrées sur les personnes, déployant les technologies numériques autour des personnes de manière humaine et dans le but principal de devenir socialement inclusives, de promouvoir la bonne gouvernance et de créer des services capables d'améliorer la qualité de vie de ses citoyens. L'objectif fondamental du Smart people est de permettre aux citoyens de devenir des acteurs essentiels de la ville intelligente [52].

5. Smart Living

La notion de vie intelligente est une composante de la ville intelligente qui couvre l'amélioration de tous les éléments de la vie quotidienne – bien-être, santé, sécurité, logement, culture, tourisme, etc. [53]. Les objectifs du Smart living sont de mettre en œuvre tous les moyens et technologies disponibles afin de faciliter, améliorer et résoudre les difficultés quotidiennes des habitants tout en augmentant l'attractivité de la ville [34]. En effet, en dehors de l'amélioration de la vie quotidienne des citoyens de la ville, elle vise également à favoriser le tourisme et les affaires en offrant un environnement attrayant.

6. Smart Governance

Les Nations unies définissent la gouvernance comme l'exercice de l'autorité économique, politique et administrative pour gérer les affaires d'un pays ou d'une ville à tous les niveaux : les mécanismes, les processus et les établissements par lesquels les citoyens et les groupes expriment leurs intérêts, exercent leurs droits légaux, respectent leurs obligations et règlent leurs litiges. Pour recentrer le citoyen dans la ville, il est essentiel d'écouter ses exigences, de construire une communication efficace avec tous les citoyens, et d'encourager l'engagement public par le développement de l'e-gouvernance pour une prise de décision participative [54].

En outre, la mise en ligne d'informations institutionnelles et la promotion de l'ouverture et de la disponibilité d'informations complètes contribuent à l'objectif de

transparence soutenu par la gouvernance intelligente. Aussi, le but est d'optimiser l'administration en développant une administration plus simple, plus rapide et plus connectée.

En plus des domaines susmentionnés, avec le développement technologique rapide, d'autres sous-domaines sont apparus, dont les suivants : Smart parking, Smart health, Smart street, Smart education, Smart home, Smart farming, Smart transport, Smart security, Smart logistic etc.

8. Le Big Data dans un Smart city

L'utilisation du Big Data dans les villes intelligentes se caractérise par un ensemble de dimensions diverses, notamment les données, la méthode de collecte des données et la valeur créée avec les données. La classification des cas existants d'utilisation des Big Data dans les villes intelligentes pourrait aider à donner un sens à cette diversité en identifiant les catégories qui partagent un certain nombre d'attributs similaires [56].

En outre, elle pourrait suggérer des modèles de référence pour la transformation des villes intelligentes basée sur les données. Dans un contexte commercial, les données peuvent provenir d'entreprises ou des clients. Par exemple, l'analyse des données peut s'appuyer sur des données de transaction commerciale pour la gestion des processus commerciaux, ou les constructeurs automobiles peuvent s'appuyer sur les données de conduite des clients. L'utilisation du Big Data dans les villes intelligentes peut également être classée selon ces mêmes perspectives, comme l'indique la littérature relative aux villes intelligentes qui traite des sources de données et des bénéficiaires [64].

Les utilisateurs de données massives dans les villes intelligentes peuvent être répartis en quatre classes.

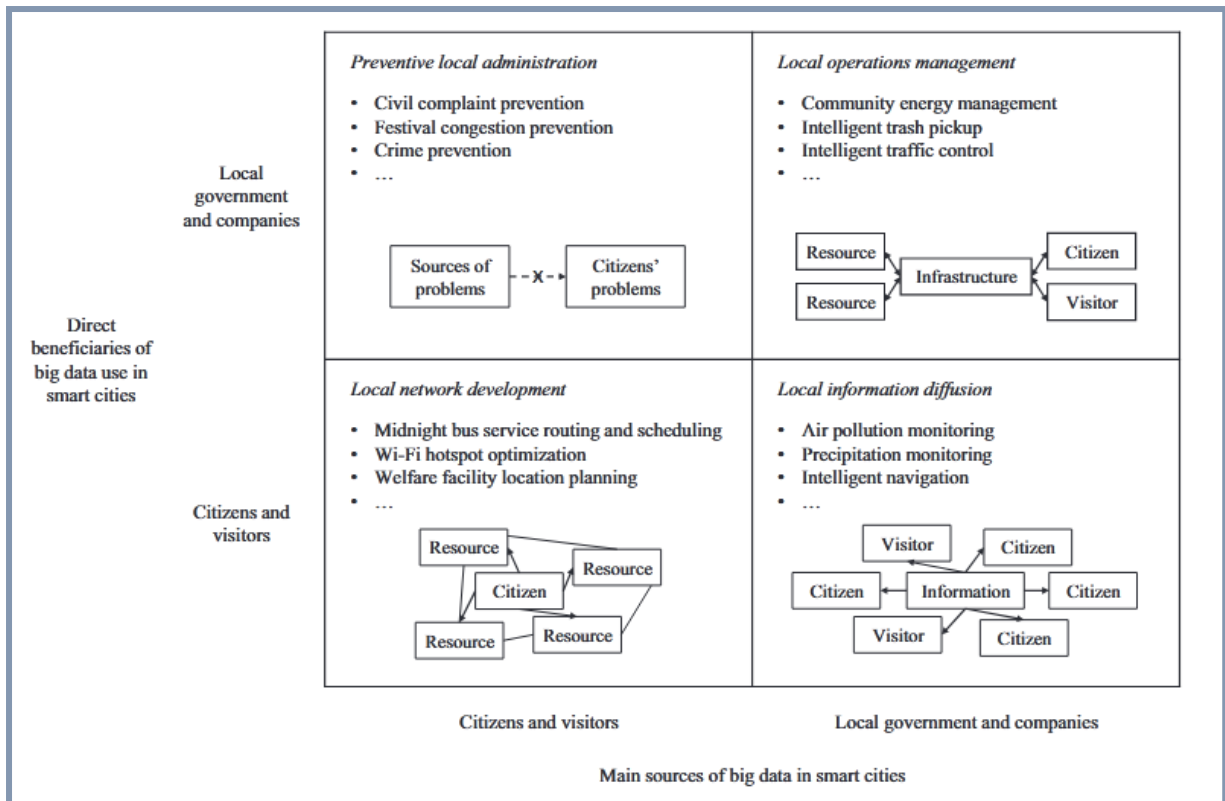


Figure 7 : Classification des cas d'utilisation du Big Data dans les villes intelligentes [56].

1. L'administration locale préventive

L'administration locale préventive (Preventive local administration) comprend des utilisations dans la prévention des crimes, des plaintes des citoyens, ou encore du trafic routier lors d'événements spécifiques (événement sportif, cérémonie, etc.)[65].

2. La gestion des opérations locales

La gestion des opérations locales (Local operations management) utilise pour sa part le Big Data dans la gestion intelligente du contrôle du trafic ou de la collecte des déchets [65].

3. Le développement des réseaux locaux

Le développement des réseaux locaux (local network development) s'appuie sur les données des clients pour aider les clients. De nombreux cas de cette catégorie se sont concentrés sur l'amélioration de l'accessibilité des réseaux locaux de ressources des clients sur la base d'une meilleure compréhension de ces derniers [66].

4. La diffusion de l'information locale

La diffusion d'informations locales (Local information diffusion) s'appuie sur les données des fournisseurs de services locaux pour aider les clients. De nombreux cas dans cette catégorie se sont concentrés sur l'identification et la diffusion efficace d'informations utiles sur les villes aux citoyens et aux visiteurs. Parmi les exemples de cette catégorie, citons la surveillance de la pollution de l'air, où les données provenant des sources de pollution de la ville ainsi que les prévisions météorologiques ont été analysées pour créer une carte de la pollution, et la navigation intelligente, où les données sur les facteurs affectant les flux de trafic ont été analysées, y compris les situations de trafic en temps réel, les accidents, la météo, la construction et les événements, à partir de capteurs placés dans la ville. Les informations identifiées ont été diffusées aux citoyens et aux visiteurs [67].

9. Défis liés à l'utilisation des Big Data dans les Smart city

Les types et les quantités de données dans les villes modernes ont explosé ces dernières années. L'omniprésence des méga données ouvre une foule de possibilités pour la transformation des villes intelligentes axée sur les données. La transition vers la transformation de la ville intelligente basée sur les données, comme toute initiative de changement à grande échelle, n'est pas facile. Les défis de la mise en œuvre de projets de villes intelligentes qui s'appuient sur l'utilisation de méga données urbaines doivent être identifiés. Dans cette partie, nous examinerons les défis qui se posent lors de la conversion des données en informations pour les villes intelligentes [56].

1. Gestion de la qualité

La qualité des données urbaines est nécessaire pour identifier des informations fiables. Pour réaliser des projets de ville intelligente qui utilisent des données, il est essentiel de prendre en compte la qualité des données disponibles, mais il est également nécessaire de découvrir des stratégies pour améliorer cette dernière au cours du projet [57].

2. Intégration des données

Les données provenant de diverses sources sont combinées. Dans les villes modernes, une variété de données est recueillie à partir de diverses sources. La clé est de relier différents

types de données afin de fournir aux citoyens et aux responsables de la ville un haut niveau de connaissances et des données de haute qualité. Cependant, étant donné que différentes organisations utilisent des structures de données différentes, il peut être difficile de connecter des données provenant de différentes sources [58].

3. Confidentialité des données

Étudier et répondre aux questions liées à des soucis de confidentialité est nécessaire afin que les projets développés créent de la valeur pour les citoyens et visiteurs de la ville [59].

4. Comprendre les besoins

Les employés des fournisseurs de services bénéficient considérablement du Big Data urbain, déterminer quelles informations sont utiles aux citoyens, aux visiteurs et aux employés est essentiel dans les projets de développement de villes intelligentes basés sur les données, car l'identification des informations qui doivent être fournies au consommateur est directement liée à la fois à la valeur d'un service et à sa demande [60].

5. Valorisation de l'information géographique

Les projets Big Data cherchent à donner des informations sur une entité géographique (rue, cite, bloc, etc.). L'acceptation des projets par les citoyens est nécessaire à leur développement [61].

6. La conception des services des Smart cities

Diverses informations sont disponibles à partir du Big Data. Le développement d'un service qui transmet ces informations peut être utile aux citoyens et aux autorités publiques. Un tel service intègre tous les résultats de l'analyse des données, de l'idéation et de la conception du contenu de l'information pour la ville intelligente [62].

Ces défis sont liés car la collecte de données, la création d'informations et la diffusion d'informations sont toutes des activités interconnectées dans les villes intelligentes basées sur les données [56].

Les six défis indiquent que l'utilisation du Big Data pour les villes intelligentes nécessite un large éventail de compétences, notamment la connaissance des citoyens,

l'administration des villes, la gestion des données, l'analyse des données, le droit et la conception des services [56].

10. Utilisation de TIC dans les Smart city

Les technologies de l'information et de la communication (TIC), influencent tous les aspects de la vie humaine : ils jouent des rôles importants sur les lieux de travail, les entreprises, l'éducation et les loisirs,

Dans d'autres secteurs, les TIC jouent aussi un rôle important. Notamment :

- Changent la façon dont les gens et les entreprises travaillent, grâce à elles ils parviennent à réduire leurs coûts mais aussi à améliorer la qualité de leurs produits [68].
- Jouent un rôle majeur dans la mise en réseau et la communication, car les entreprises les utilisent pour faciliter la communication entre les employés [69].
- Apportent un soutien dans des domaines tels que la collecte et la diffusion d'informations [68].
- Facilitent grandement l'acquisition et l'absorption des connaissances [70].

Ce point a d'abord montré que les TIC sont des moyens de résoudre les problèmes publics auxquels les villes intelligentes sont confrontées. Deuxièmement, qu'elles sont identifiées dans de nombreux exemples comme des catalyseurs de changement.

Par conséquent, les TIC sont des moyens de résoudre les problèmes publics auxquels les villes intelligentes sont confrontées. Ensuite, elles sont identifiées dans de nombreux exemples comme des facteurs de changement [71].

Les infrastructures nécessaires au déploiement des villes intelligentes sont les capteurs et les équipements numériques (TIC). Ces objets permettent de partager des informations en temps réel [71].

Le premier défi consiste à promouvoir la numérisation des services de la ville, en pensant du point de vue de l'utilisateur [71].

Le deuxième défi concerne la favorisation de la numérisation des services de la ville, en se plaçant du point de vue de l'utilisateur. Il ne s'agit pas seulement d'un objectif de l'administration publique, mais aussi de l'industrie du logiciel, qui souhaite attirer les utilisateurs sur ses plates-formes [71].

11. Conclusion

Le principal objectif d'une ville intelligente est de réduire les coûts engendrés par la mise à disposition des services publics et de répondre au mieux aux exigences des utilisateurs tout en assurant :

- Une réduction des charges de fonctionnement de la ville grâce à des capteurs permettant d'optimiser la maintenance des équipements ou au remplacement de certains matériels par des systèmes moins énergivores.
- Une automatisation de certaines tâches afin de leur permettre de gagner en productivité et de consacrer du temps sur d'autres actions.
- Le développement de nouveaux services ou encore la simplification de services existants.

CHAPITRE III : PROPOSITION ET IMPLEMENTATION D'UN NOUVEAU MODELE DE STOCKAGE DE BIG DATA

1. Introduction

Dans ce chapitre, nous signalons les faiblesses des solutions existantes du stockage de données massives. Puis, nous proposons un nouveau modèle basé sur le principe de partage des ressources du stockage de données au niveau des nœuds IoT afin de mettre un nouveau modèle de stockage distribué.

Dans le projet que nous allons mettre en place, nous nous sommes appuyés sur un simulateur des villes intelligentes afin de trouver des solutions tangibles et secondaires. Cela permet de réduire le taux de perte de données lors les échange entre les émetteurs et récepteurs.

2. Le stockage de données dans les villes intelligentes

Tandis que les villes s'appuient depuis longtemps sur la vidéosurveillance pour améliorer la sécurité publique et prévenir la criminalité, les villes intelligentes d'aujourd'hui exploitent ces solutions pour répondre à une multitude d'autres besoins, notamment la gestion du trafic, l'éclairage et le contrôle du stationnement.

Le volume cumulé de données générées dans le monde entier au cours des dernières années a augmenté à un rythme très rapide, une grande partie de ces données provenant des caméras de surveillance. Parmi Les solutions au problème du stockage des données de vidéosurveillance comprennent la préservation des données à la périphérie pour des analyses immédiates et des renseignements en temps réel [72].

À l'ère de l'Internet des objets (IoT) et de l'intelligence artificielle (IA), où les appareils génèrent plus de données que jamais, trouver un moyen de stocker convenablement toutes ces données est devenu une nécessité. En fonction des besoins d'une ville intelligente, plusieurs approches sont envisageables, notamment le stockage à la périphérie (près de l'emplacement), dans le Fog, dans le cloud ou sur des serveurs spécifiques [73].

1. Solutions existantes

Une grande variété de dispositifs IoT, utilisés pour exécuter des applications utilisateurs, sont répartis sur une large zone géographique [75]. Par conséquent, un très grand volume de Big Data doit être stocké sur des différentes ressources de stockage tels que : (1) le Edge, (2) le Fog ou (3) le cloud. Le contrôle de l'endroit où les données sont stockées devient aussi important que le contrôle de ces dernières [75]. Par conséquence, plusieurs questions se révèlent, comme (1) où placer les données ? car l'emplacement des données est un facteur important pour un calcul efficace et rapide ; (2) comment gérer une quantité massive de demandes de transfert de données ? et (3) comment gérer les ressources de stockage.

Aussi, dernièrement plusieurs travaux ont démontré l'importance du stockage distribué sur les différents niveaux de nouvelles architectures tels que [75], [76], [77],[78],[79].

2. Discussion

Le cloud est certainement le meilleur endroit pour stocker et traiter les données, car il dispose de ressources de stockage et de traitement très élevées. Dans ce cas, les données sont traitées et transmises via Internet et non sur un disque dur ou un serveur local, il permet également aux utilisateurs d'accéder aux documents depuis n'importe quel endroit, tant qu'ils ont accès à Internet[74].Cependant, en raison de l'éloignement des data centers dans le cloud et de la congestion du trafic réseau, cette solution offre souvent des performances limitées, principalement en terme de la latence de l'utilisateur final. En outre, l'envoi immédiat de toutes les données vers le cloud dégrade considérablement les performances de cette solution [75].

Donc, le stockage au niveau du cloud présente certaines limites, dont :

- La limite de temps de traitement des données en temps réel. En effet, les données collectées à partir des machines et des capteurs sont envoyées vers le cloud et les sorties souhaitées sont renvoyées aux appareils souhaités, ce qui peut ralentir la réponse [80].
- Les données massives nécessitent une bande passante élevée, ce qui est coûteux.
- L'indisponibilité du réseau, ces problèmes deviennent des barrières applicatives en raison de la limite de temps de requête.
- La diffusion en continu de grandes quantités de données dans le cloud peut présenter des problèmes importants de latence utilisateur.

Concernant le stockage au niveau des Fog [75], cette solution temporaire permet de stocker les données au plus près de l'endroit où elles sont collectées. Toutefois, les données sont transmises du point de collecte à une passerelle pour y être traitées, puis renvoyées à la périphérie. Il agit comme un pont, réunissant le cloud et la périphérie.

Cependant, le stockage au niveau du Fog présente certaines limites, dont :

- Les Fog possèdent des ressources de stockage limitées ; ce qui nécessite le transfert des données vers le cloud une fois l'espace mémoire est insuffisant pour contenir les nouvelles données en entrée.
- Le partitionnement du réseau à cause de la mobilité des dispositifs IoT.
- Le problème de la latence utilisateurs élevée est toujours présent, à cause de l'éloignement des dispositifs IoT du Fog.
- La congestion du trafic réseau.

En fin, le stockage au niveau de l'Edge minimise considérablement la latence utilisateur mais pour une courte durée du temps qui précède la saturation de l'espace mémoire Edge. Après cela, cette solution offre des performances limitées à cause de :

- Espace mémoire au niveau de l'Edge est très limité par rapport au cloud ; ce qui nécessite le transfert des données vers le cloud ou le Fog.
- Le partitionnement du réseau.
- La congestion réseau et la consommation de la bande passade.
- Etc.

3. Un algorithme de stockage distribué des données

Notre objectif est de proposer un modèle de stockage de données distribué (figure 8) afin d'exploiter toutes les ressources de stockage disponibles dans les dispositifs IoT et qui sont considérés comme le niveau bas de l'architecture de l'internet des objets en réduisant le nombre d'accès au cloud et en maintenant les informations à proximité de l'utilisateur final tout en fournissant un stockage distribué basé sur la collaboration des appareils IoT. Cette solution peut fournir un espace de stockage qui dépasse largement l'espace de l'Edge. Bien sûr, une copie de ces données sera périodiquement copiée vers le cloud pour un stockage permanent, puisque les ressources de stockage des appareils IoT ne fournissent qu'un stockage temporaire des données.

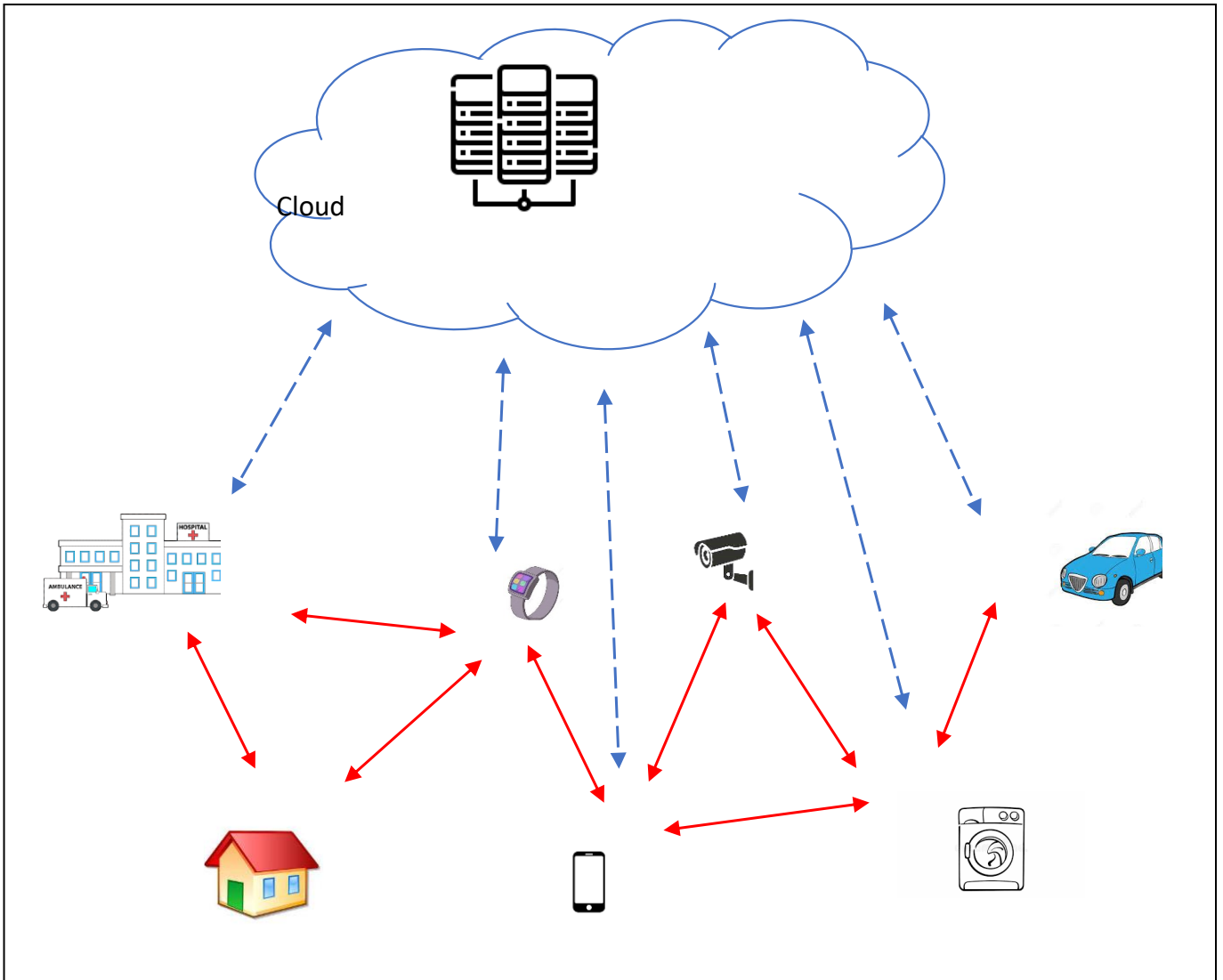


Figure 8 : modèle de stockage de données distribué.

1. Description de l'algorithme proposé

L'algorithme proposé est composé principalement de deux fonctions suivantes :

- `Partage_de_ressource_de_stockage()` : cette fonction permet de répondre positivement à une demande de stockage distribué si l'espace mémoire local du récepteur est suffisant, sinon le récepteur ignore tout simplement cette dernière (algorithme1). Cela permet de réduire le nombre de message générée et donc, d'améliorer sa complexité.
- `Stockage_distribué ()` : cette fonction permet de recevoir les données collectées et les stocker dans le cas où la taille de l'espace mémoire de l'appareil IoT est suffisant pour contenir ces données. Sinon, une demande d'autorisation du stockage distribué sur un autre dispositif de son réseau est diffusée (algorithme 2). L'émetteur de la demande attend la réception d'un message d'autorisation pour une certaine durée du temps avant de rediffuser à nouveau une autre demande.

Variables :

Id : identificateur du processus courant initialisé à p ;

id_dem : processus initialisé à Null ;

Id : identificateur du processus courant initialisé à p ;

Récepteur, idR : des processus initialisés à Null.

Seuil : entier définie par l'utilisateur.

Espace mémoire local libre : entier initialisé à la taille de l'espace mémoire utilisateur.

Réponse : booléen initialisé à faux.

Algorithme 1 :Partage_de_ressource_de_stockage()

Début

Recevoir (demande, id_dem)

Si (espace_mémoire_local_libre > seuil) alors

Envoyer (autorisation, id) ;

Fin.

Sachant que *P* est le processus courant.

Algorithme 2 :Stockage_distribué ()

Début

Réception(données)

Si (espace_mémoire_local_libre > seuil) alors

Stocker (data, mémoire)

Sinon

Stocker(data, espace_de_reserve) ;

Diffuser(demande, Id) ;

Libérer espace local(cloud) ;

Tant que(réponse==faux) faire

Début

Si Recevoir (autorisation, idR) alors

Récepteur =idR ;

Envoyer (data, récepteur) ;

Sinon

Diffuser (demande, Id) ;

Fin tant que ;

Fin.

2. Les avantages de la solution proposée

Parmi les avantages de la solution proposée pour le stockage de données massives dans une ville Smart, nous citons les points suivants :

- Réponse en temps réel.
- Réduction du trafic réseau.
- Réduire l'utilisation de la bande passante.
- Permet d'atteindre et d'éviter les encombrements.
- Améliorer la qualité de service (QoS) dans ce type de système.
- Minimiser le nombre d'accès vers le cloud tout en profitant de la collaboration de différents dispositifs IoT.
- Fournir plus d'espace de stockage qui dépasse de loin l'espace d'un seul appareil.
- Une puissance de calcul importante pour optimiser le traitement des données et fournir de meilleures performances, en répartissant les charges de travail entre la périphérie et le cloud.
- Les notifications transmises dans les meilleurs délais.

4. Exemple du modèle proposé

Afin de clarifier le modèle proposé pour le stockage des données dans la ville intelligente, nous avons présenté un modèle proche de la réalité avec deux émetteurs (capteur de gaz, capteur de température), plusieurs routeurs pour le flux de données et quatre récepteurs des données générées pour la réception, avec la précision des routes (chemin) à suivre.

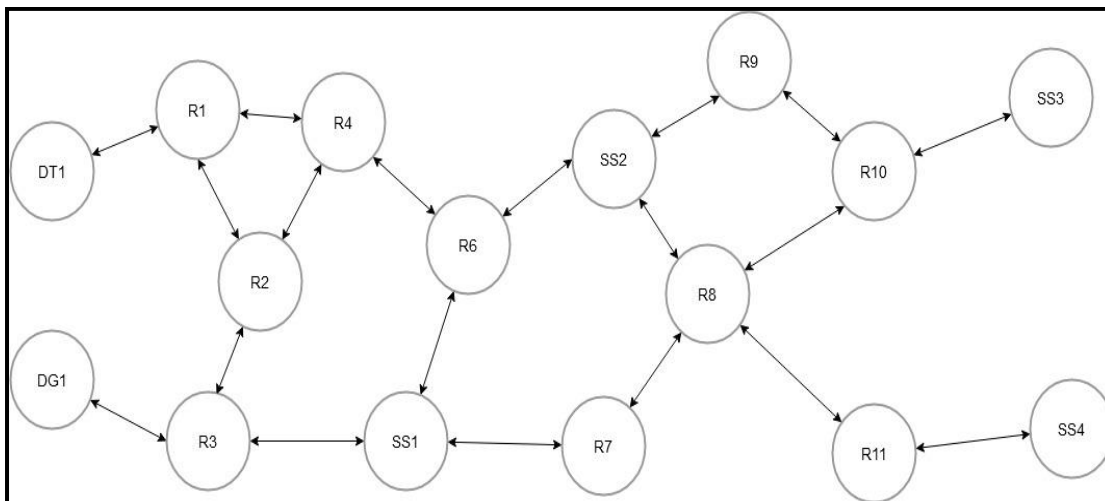


Figure 9 : Le modèle de stockage distribué de Big Data

Dans ce modèle, nous avons utilisé les nœuds IoT énumérés dans le tableau 1 :

Abréviation	Signification
DG1	Détecteur de gaz
DT1	Détecteur de température
R1...R11	Routeur
SS1...SS4	Support de stockage des données

Tableau 1 : Abréviations utilisées pour définir les nœuds IoT.

5. Environnement de développement

Nous avons utilisé le logiciel « CupCarbon » parce que c'est un simulateur de ville intelligente, gratuit, simple et facile à représenter les capteurs, les nœuds, les stations de base et aussi la mobilité.

1. Représentation de CupCarbon

Le CupCarbon [81] est un simulateur de réseau de capteurs sans fil (SCI-WSN) pour les villes intelligentes et l'Internet des objets. Son objectif est de créer des scénarios environnementaux tels que des incendies, du gaz, des mobiles, et plus largement à l'intérieur d'initiatives éducatives et scientifiques en concevant, visualisant, déboguant et validant des algorithmes distribués pour la surveillance, la collecte de données environnementales, etc.

Il peut non seulement servir à illustrer visuellement les concepts fondamentaux des réseaux de capteurs et leur fonctionnement, mais il peut aussi aider les scientifiques à tester leurs topologies sans fil.

CupCarbon fournit deux environnements de simulation. Le premier environnement de simulation permet de créer des scénarios de mobilité et de réaliser des phénomènes naturels tels que des incendies et des gaz, ainsi que de simuler des mobiles tels que des automobiles et des objets volants (par exemple, des drones, des insectes, etc.).

Le second environnement de simulation représente une simulation à événements discrets de réseaux de capteurs sans fil qui prend en compte le scénario construit sur la base du premier environnement.

Comme le présente la Figure 9, l'interface utilisateur graphique (GUI) de CupCarbon se compose des cinq parties principales suivantes :

1. La carte (au centre)
2. La barre de menu (en haut)
3. La barre d'outils (sous le menu)
4. Le menu des paramètres (à gauche)
5. La barre d'état (en bas)
6. La console

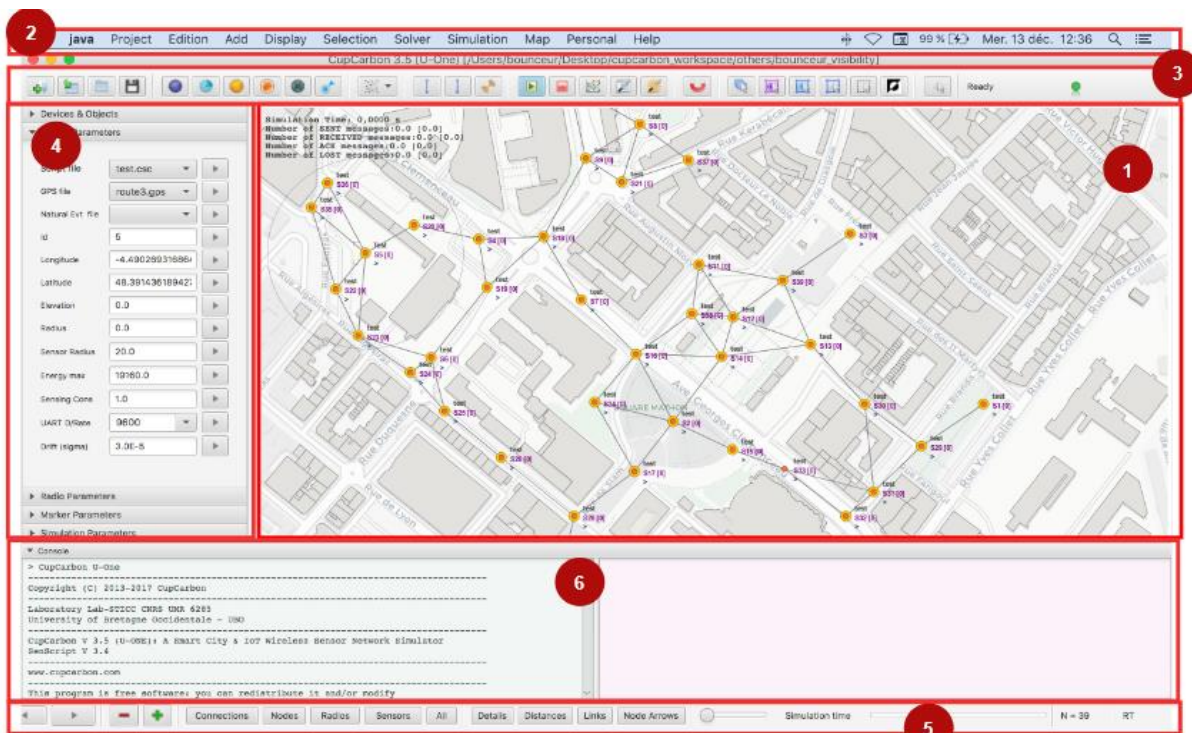


Figure 10 : Interface utilisateur de CupCarbon [81].

2. La carte

La carte est l'objet principal du simulateur CupCarbon. C'est la partie où le réseau et les objets du projet peuvent être conçus. La carte peut être modifiée en fonction des préférences de l'utilisateur ou de la manière dont les informations doivent être présentées. Le Tableau 2 montre les 12 cartes ou arrière-plans possibles, qui sont :

Simulateur	Partie	Arrière-plans
CupCarbon	La carte	OpenStreetMap clair
		OpenStreetMap Fond sombre
		Fond en pointillés

	Fond quadrillé blanc à petites cellules
	Fond quadrillé noir à petites cellules
	Fond noir
	Fond blanc
	Fond de cellules grises moyennes
	Arrière-plan du carnet
	Fond de cellules blues moyennes
	Carte Google
	Carte Google (satellite)

Tableau 2 : Les différents arrière-plans de la carte du simulateur CupCarbon.

3. La barre de menu

Comme le montre la Figure 10, la barre de menu est composée de 10 éléments qui sont :

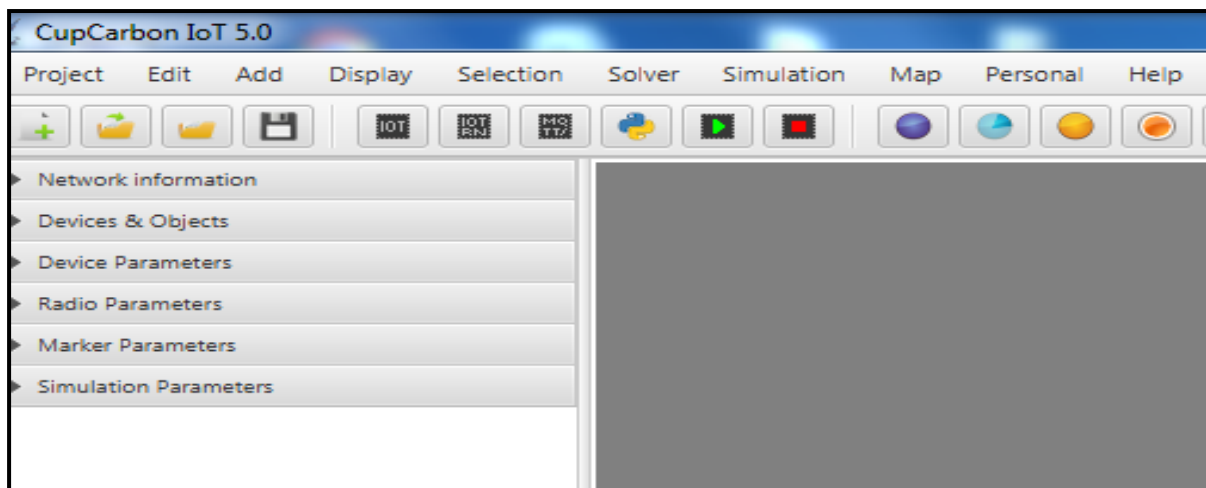


Figure 11 : La barre menu de CupCarbon [81].

Le tableau 3 contient les 10 composants de la barre de menu et leurs fonctionnements.

Simulateur	Partie	Composants	Fonctionnements
CupCarbon	Barre de Menu	Project	<ul style="list-style-type: none"> - Nouveau projet. - Nouveau projet à partir du projet actuel. - Ouvrir un projet. - Ouvrir le dernier projet. - Projets récents. - Réinitialiser. - Quitter.

		Edition	<ul style="list-style-type: none"> - Annuler. - Refaire. - Dupliquer. - Supprimer.
CupCarbon	Barre de Menu	Add	<ul style="list-style-type: none"> - Ajout de nœuds de capteurs. - ajout de nœuds de capteurs directionnels. - Ajout de stations de base. - Ajout de gaz. - Ajout de mobiles. - Ajout de marqueurs. - Ajout de nœuds météorologiques. - Ajout de nœuds de capteurs aléatoires.
		Display	<ul style="list-style-type: none"> - Afficher/Masquer les détails. - Afficher/Masquer le nom du fichier. -Afficher/Masquer les niveaux de batterie/tampon. - Afficher/Masquer les distances radio. - Afficher/Masquer les messages radio. -Afficher/Masquer les distances des marqueurs. - Afficher/Masquer les liens. - Afficher/Masquer les flèches du réseau. -Afficher/Masquer les flèches des marqueurs. - Afficher/Masquer les bâtiments. - Couleur du lien suivant et Couleur du lien précédent.
		Sélection	<ul style="list-style-type: none"> - Sélectionner tout. - Désélectionner tout. - Inverser la sélection. - Sélectionner les capteurs sans script. - Sélectionner les capteurs sans GPS. - Sélectionner les capteurs de marqueurs. - Ajouter une sélection Sélectionner toutet désélectionner tout.

		Simulation	<ul style="list-style-type: none"> - Exécution de la simulation. - Arrêt de la simulation. - consommation d'énergie. - Exécuter la simulation. - Arrêter la simulation. - Consommation d'énergie. - Fenêtre SenScript. - Générateur d'événements naturels. - Tester la mobilité et les événements sélectionnés. - Tester toutes les mobilités et tous les événements. - Tester les mobilités des nœuds de capteurs. - Tester les mobilités des mobiles. - Arrêter de tester les mobilités et les événements. - Initialiser tout.
CupCarbon	Barre de Menu	Solver	<ul style="list-style-type: none"> - Algorithme de couverture des capteurs. - Algorithme de couverture de la cible. - Algorithme de couverture de la cible (Th). - Ordonnancement. - Centre de réseau. - Coque.
		Map	Voir tableau 2 pour plus d'informations
		Personal	Le menu est utilisé pour créer des fonctions SenScript personnelles
		Help	Le menu Aide contient le lien d'aide et la boîte à propos.

Tableau 3 : La partie barre de menu du simulateur CupCarbon.

4. La barre d'outils

La barre d'outils de CupCarbon est représentée dans la Figure 11 et permet d'accéder aux principales actions de CupCarbon rapidement.

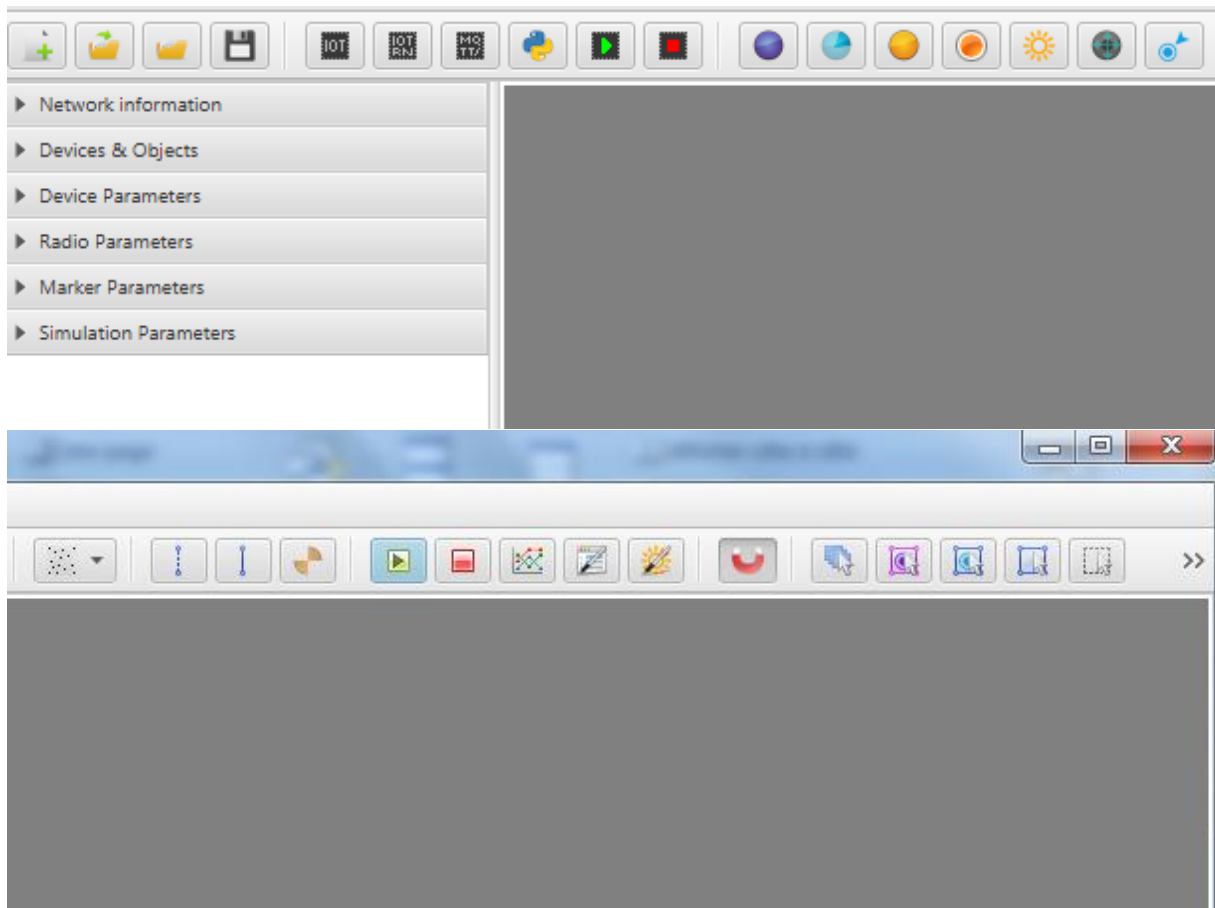


Figure 12 : La barre d'outils de CupCarbon [81].

5. Le menu des paramètres

Le menu des paramètres est une partie essentielle dans le simulateur car il donne une perspective générale du travail. Le tableau 4 présente les 6 paramètres de base

Simulateur	Partie	Composants	Fonctionnements
CupCarbon	Menu des paramètres	Panneau d'information sur le réseau	Affiche certaines informations sur le réseau, comme le nombre de capteurs, le nombre de capteurs marqués, etc.
		Périphériques et objets	-sélectionner un objet sur la carte par son nom -sélectionner/désélectionner des objets par leur type
		Panneau Paramètres du dispositif	Ce panneau permet de modifier les paramètres des objets sélectionnés comme : Fichierscript, Fichier GPS, Fichier d'événements naturels, Id, etc..

		Panneau des paramètres radio	Ce panneau permet de modifier les paramètres du module radio des nœuds de capteurs sélectionnés
		Panneau Paramètres des traceurs	Utiliser pour définir une route entre deux points
		Paramètres de simulation et panneau SenScript	Permet de démarrer ou d'arrêter une simulation et de choisir la vitesse de la simulation

Tableau 4 : La partie menu des paramètres du simulateur CupCarbon.

Dans le composants Panneau Paramètres du dispositif, Toute modification n'est prise en compte que si elle est suivie d'un clic sur le bouton d'application avec flèche grise droite situé dans la partie droite du champ correspondant.

6. La barre d'état

La barre d'état est située en bas de la fenêtre principale de CupCarbon. Elle permet d'afficher certaines informations et contient des boutons permettant de faire certaines options comme dans le cas de la barre d'outils.

7. La console

La console est utilisée par le simulateur pour afficher certains messages utiles à l'utilisateur pendant la simulation. Elle est composée de deux parties. La première, à gauche, est utilisée pour afficher les messages concernant la simulation. Il est également possible d'afficher les messages des capteurs en utilisant la commande SenScript cprint. La deuxième partie à droite est utilisée pour afficher les erreurs pendant la simulation. Dans les nouvelles versions de CupCarbon, cette fenêtre de la console est séparée de l'interface principale.

8. CupCarbon Objects

Cette section contient une liste de tous les objets qui peuvent être placés sur la carte.

Nœud de capteur : est l'objet principal de CupCarbon ! Il contient des pièces principales : des modules radio, une unité de détection et une batterie. Un nœud capteur peut être ajouté en cliquant sur le menu Ajouter.

Nœud de capteur directionnel : est le même que le nœud de capteur classique avec un autre type d'unité de détection, qui est directionnelle. Ce dernier n'est pas circulaire, il a une forme de cône.

Station de Base (Sink) : La station de base est exactement le même objet que le nœud capteur à l'exception qu'il a une batterie infinie.

Gaz (événements analogiques) : Le gaz ou l'événement naturel permet de générer des valeurs analogiques. Il peut être mobile. Son objectif est de simuler des valeurs aléatoires ou données provenant de l'environnement.

Mobile : Un mobile dépend de son itinéraire. Ce dernier est créé à l'aide de deux marqueurs, il suffit de sélectionner l'itinéraire qui sera suivi par le mobile dans la liste des itinéraires créés depuis le champ Fichier GPS dans la vue Paramètres de l'appareil.

Marker : Les marqueurs peuvent être utilisés pour de nombreux objectifs.

Weather : Le module météo permet de générer n'importe quel paramètre météo et notamment la température.

9. Les langages utilisés

10. Python

Python est un langage de programmation interprété multiparadigme et multi-plateforme. Il encourage une programmation impérative structurée, fonctionnelle et orientée objet.

11. SenScript

SenScript [81] est le script utilisé pour programmer les nœuds de capteur du simulateur CupCarbon. Il est simple à utiliser et très similaire à un langage de programmation distribué naturel. Python peut également être utilisé pour programmer les nœuds, leur donnant plus de flexibilité et de possibilités.

Exemple :

SenScript : set x abcd à Java : String x = "abcd » ;

SenScript : set y x à Java : String y = "x » ;

SenScript : set y \$x à Java : String y = x ;

Un autre exemple de code SenScript est illustré dans le script suivant :

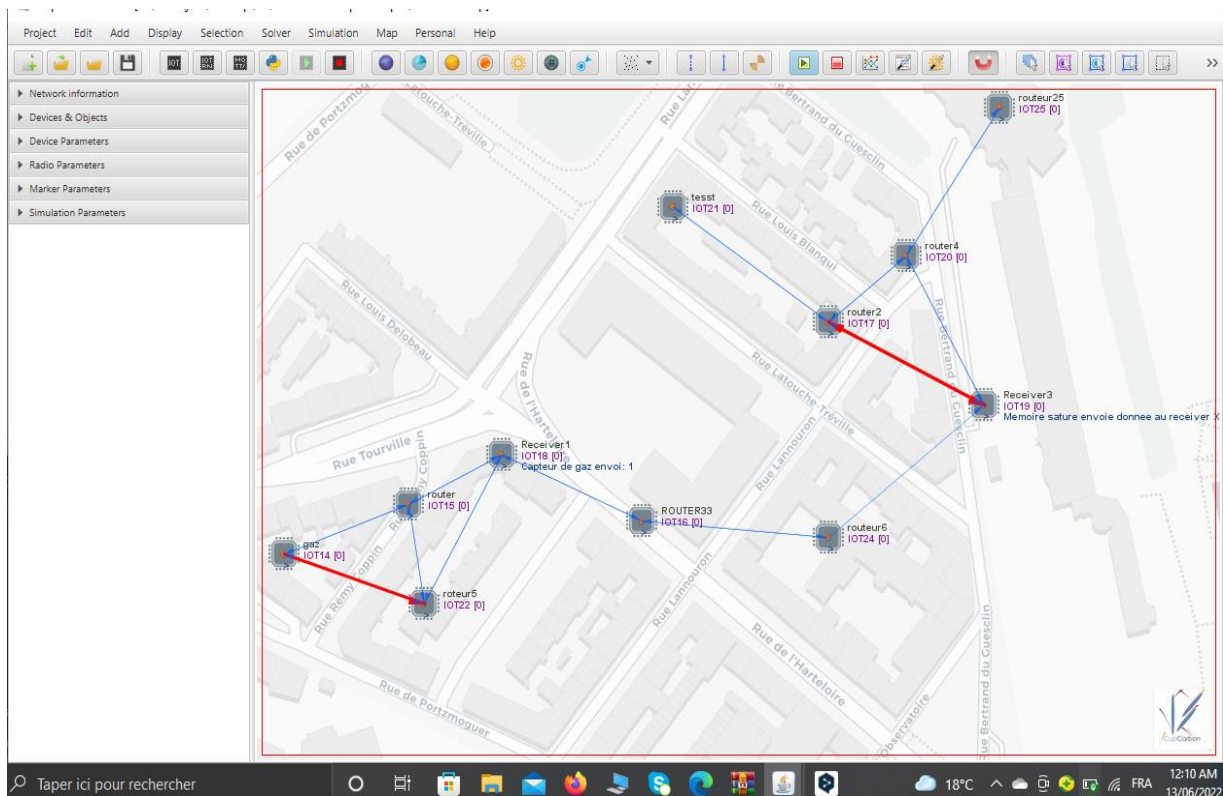


Figure 14 : génération des données par les capteurs et les envoyer vers les routeurs.

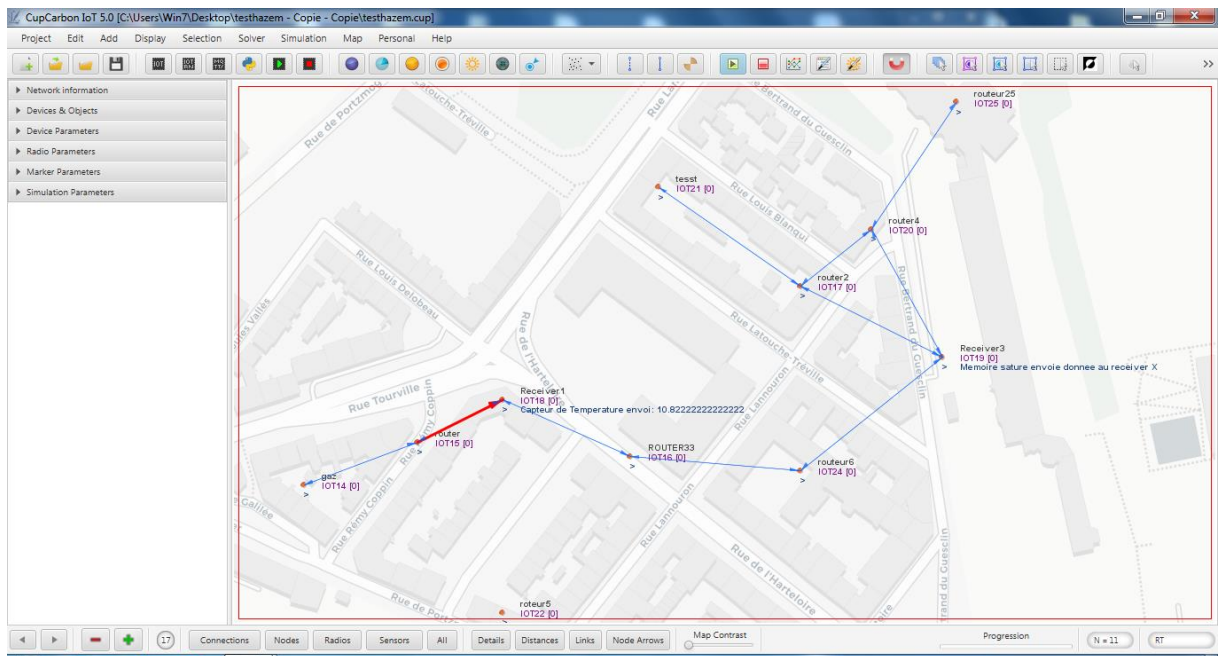


Figure 15 : l'envoi de données par les capteurs vers les récepteurs appropriés au travers les routeurs.

Étape 02 : Après un certain temps, la mémoire de récepteur sera saturée, comme solution il transmet les données vers un autre récepteur. La figure suivante représente le résultat de cette étape :

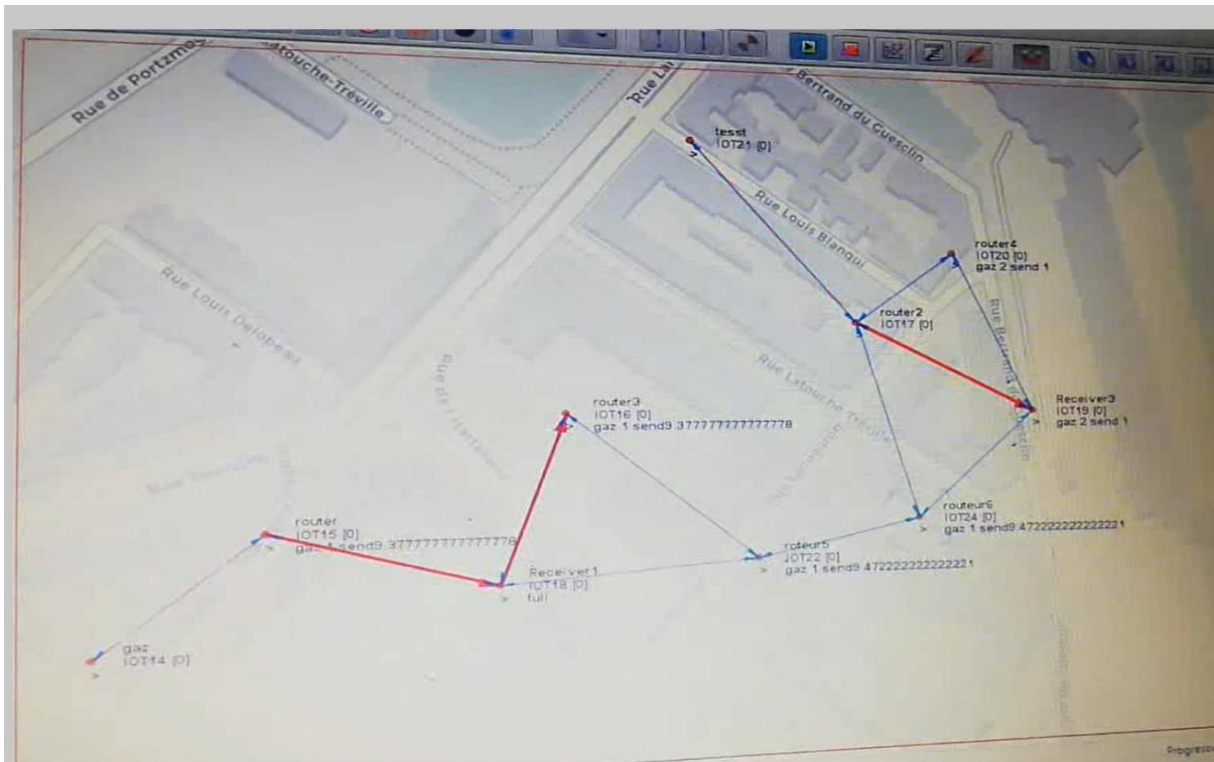


Figure 16 : redirection des requêtes des capteurs vers un récepteur distant pour le stockage distribué (1).

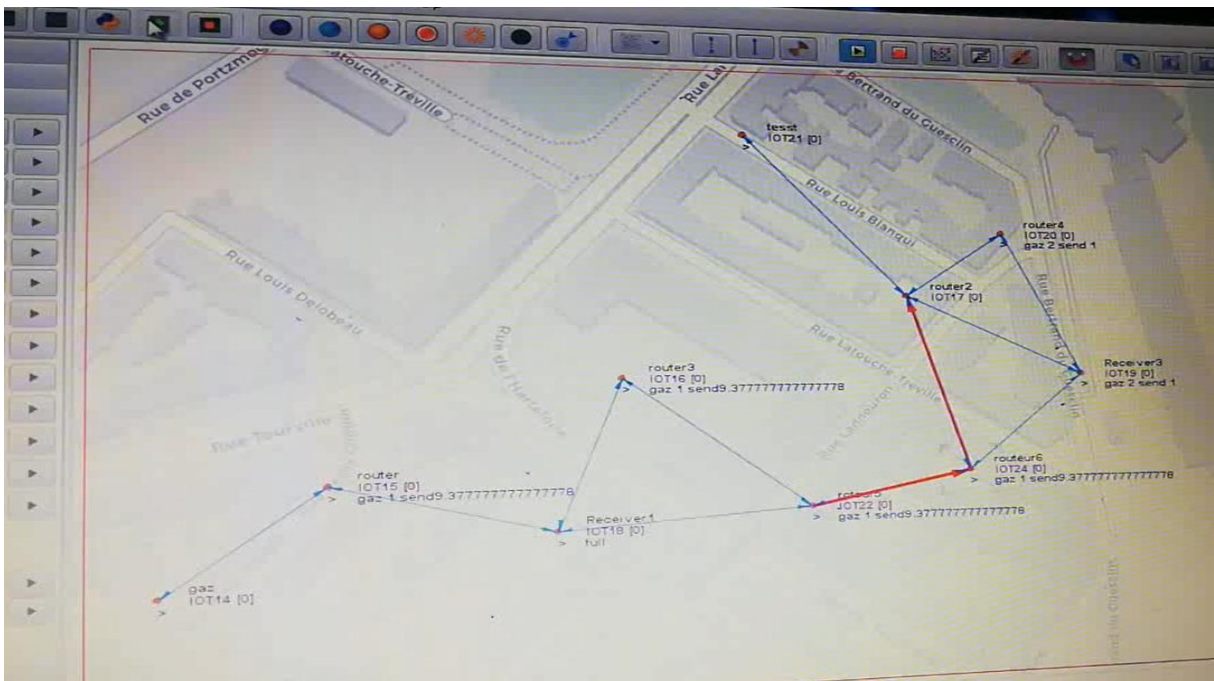


Figure 17 : redirection des requêtes des capteurs vers un récepteur distant pour le stockage distribué (1).

Étape 3 : Après la libération de l'espace mémoire de nœud émetteur en envoyant les données enregistrées vers le cloud par vague, il recommence à enregistrer les données sur sa mémoire locale. Cette étape se répète durant toute la durée de fonctionnement des capteurs.

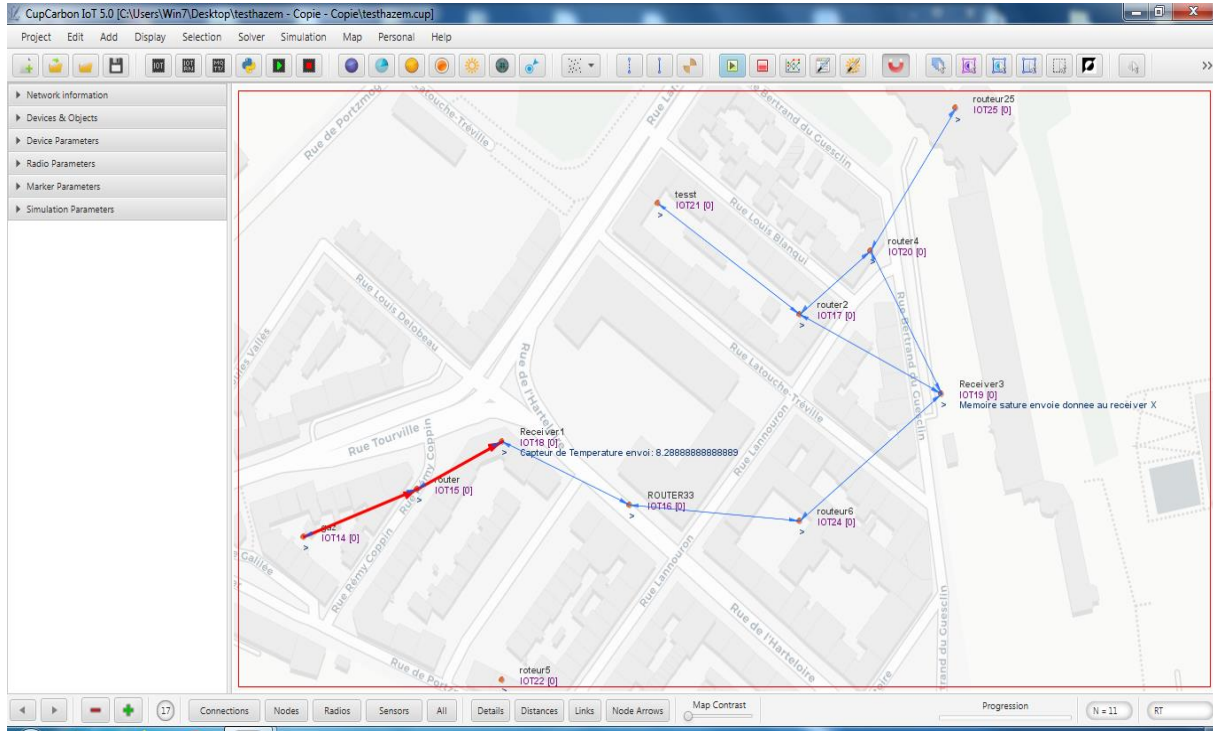


Figure 18 : libération de la mémoire et enregistrement des données sur l'espace interne.

7. Conclusion

Dans ce chapitre, nous avons présenté une description détaillée de la solution proposée ainsi que les avantages de cette contribution. Ensuite, nous avons présenté le simulateur de villes intelligentes utilisé ainsi que les résultats d'implémentation de notre idée.

CONCLUSION GENERALE

Le concept de ville intelligente est un domaine multidisciplinaire où un système devient très complexe en raison de ces nombreux secteurs et très volumineux en raison de l'énorme quantité de données et de connaissances nécessaires à sa fonctionnalité. L'évolutivité dans ce domaine est donc fondamentale, une ville doit développer un système qui progresse avec son développement.

Parmi les domaines liés à la ville intelligente, qui connaît un développement et une explosion importants, on trouve le Big Data qui est directement en relation avec les caractéristiques et les exigences de la ville intelligente.

L'un des défis du Big Data est le stockage des données et la rapidité d'accès à celles-ci afin de les traiter et de les analyser dans un délai court et d'en tirer profit.

Dans ce projet et vu la problématique posée on a proposé une nouvelle méthode de stockage de données distribué qui collecte les données aux périphériques de réseau et les traite en temps réels et qu'offre une faible latence.

En perspective, ce travail peut être complété par les points suivants :

1. Finaliser l'implémentation et l'évaluation de performances de notre contribution.
2. Tester ce projet à l'échelle réelle.
3. Intégrer des nouvelles techniques de mobilité et de partage de ressources.

REFERENCE

- [1] Bahga, Arshdeep, and Vijay Madisetti. Big Data science & analytics: A hands-on approach. VPT, 2016.
- [2] Beyer, Mark A., and Douglas Laney. "The importance of 'Big Data': a definition." Stamford, CT: Gartner (2012): 2014-2018.
- [3] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448
- [4] Chen, Y., Chen, H., Gorkhali, A., Lu, Y., Ma, Y., & Li, L. (2016). Big Data analytics and Big Data science: a survey. *Journal of Management Analytics*, 3(1), 1-42.
- [5] Clavert, Frédéric. "Patrick Manning, Big Data In History". Lectures, 2014. Openedition,
- [6] Futura, and @futurasciences. "Définition | Internet Des Objets - IdO - Internet of Things – IOTFutura Tech." *Futura*, www.futura-sciences.com, <https://www.futura-sciences.com/tech/definitions/internet-internet-objets-15158/>. Dernier accès le 1 June 2022.
- [7] Hurwitz, Judith S., et al. Big Data for dummies. John Wiley & Sons, 2013.
- [8] Eberendu, Adanna Cecilia. "Unstructured Data: an overview of the data of Big
- [9] Feldman, Ronen, and James Sanger. The text mining handbook: advanced approaches in
- [10] Hänig, Christian, Martin Schierle, and Daniel Trabold. "Comparison of structured vs. unstructured data for industrial quality analysis." Proceedings of The World Congress on Engineering and Computer Science. 2010.
- [11] N.benhamida, cours analyse de données massives, 2eme Master STIC. Année 2021-2022.
- [12] "What Is Big Data? - Types, Advantages, and Characteristics - Intellipaat." *Intellipaat Blog*, intellipaat.com, 9 Dec. 2019, <https://intellipaat.com/blog/tutorial/hadoop-tutorial/big-data-overview/>. Dernier accès le 1 June 2022.
- [13] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big Data concepts, methods, and analytics. *International journal of information management*, 35(2), 137-144.

- [14] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.
- [15] Demchenko, Yuri, Cees De Laat, and Peter Membrey. "Defining architecture components of the Big Data Ecosystem." 2014 International Conference on Collaboration Technologies and Systems (CTS). IEEE, 2014.
- [16] JOUANNOT, O. (2013). Présentation Générale Big Data. *Guide Share France*.
- [17] Excelsior. "Big Data, Explained: The 5V s of Data | by Excelsior | Medium." *Medium*, medium.com, 30 Jan. 2022, https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1. Dernier accès le 1 June 2022.
- [18] Chen, M., Mao, S., & Liu, Y. (2014). Big Data: A survey. *Mobile networks and applications*, 19(2), 171-209.
- [19] Lenzerini M (2002) Data integration: a theoretical perspective. In: Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on principles of database systems. ACM, pp 233–246.
- [20] Maletic JI, Marcus A (2000) Data cleansing: beyond integrity analysis. In: IQ. Citeseer, pp 200–209.
- [21] Brewer EA (2000) Towards robust distributed systems. In: PODC. p 7
- [22] Gilbert S, Lynch N (2002) Brewer’s conjecture and the feasibility of consistent, available, partition-tolerant web services. *ACM SIGACT News* 33(2):51–59
- [23] Chaiken R, Jenkins B, Larson P-A°, Ramsey B, Shakib D, Weaver S, Zhou J (2008) Scope: easy and efficient parallel processing of massive data sets. *Proc VLDB Endowment* 1(2):1265– 1276
- [24] Beaver D, Kumar S, Li HC, Sobel J, Vajgel P et al (2010) Finding a needle in haystack: facebook’s photo storage. In OSDI, vol 10. pp 1–8
- [25] DeCandia G, Hastorun D, Jampani M, Kakulapati G, Lakshman A, Pilchin A, Sivasubramanian S, Vosshall P, Vogels W (2007)

- Dynamo: amazon's highly available key-value store. In: SOSP, vol 7. pp 205–220
- [26] Chang F, Dean J, Ghemawat S, Hsieh WC, Wallach DA, Burrows M, Chandra T, Fikes A, Gruber RE (2008) Bigtable: a distributed storage system for structured data. *ACM Trans Comput Syst (TOCS)* 26(2):4
- [27] Chodorow K (2013) MongoDB: the definitive guide. O'Reilly Media Inc
- [28] Labrinidis A, Jagadish HV (2012) Challenges and opportunities with Big Data. *Proc VLDB Endowment* 5(12):2032–2033.
- [29] Sinha, Shubham. “Real Time Big Data Applications in Various Domains | Edureka.” *Edureka*, www.edureka.co, 29 Jan. 2018, <https://www.edureka.co/blog/big-data-applications-revolutionizing-various-domains/>. Dernier accès le 3 June 2022.
- [30] Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of urban technology*, 22(1), 3-21.
- [31] Marsal-Llacuna, J. Colomer-Llina's, and J. Mele'ndez-Frigola, “Lessons in urban monitoring taken from sustainable and livable cities to better address the Smart Cities initiative, Technological Fore-casting and Social Change” (2014).
- [32] M. O'Grady, and G. O'Hare, “How Smart Is Your City?” *Science* 335: 3 (2012) 1581 – 1582.
- [33] Harrison, C., Eckman, B., Hamilton, R., Hartswick, P., Kalagnanam, J., Paraszczak, J., & Williams, P. (2010). Foundations for Smarter cities. *IBM Journal of research and development*, 54(4), 1-16
- [34] Breux, S., & Diaz, J. (2017). La ville intelligente : origine, définitions, forces et limites d'une expression polysémique.
- [35] Nam, T., & Pardo, T. A. (2011, June). Conceptualizing Smart city with dimensions of technology, people, and institutions. In *Proceedings of the 12th annual international digital government research conference: digital government innovation in challenging times* (pp. 282-291).
- [36] Sajhau, P. (2017). IBM–Building sustainable cities through partnerships and integrated approaches. *Field Actions Science Reports. The journal of field actions*, (Special Issue 16), 52-57.

- [37] “Smart city : Quand l’innovation Relève Les Défis de La Ville Durable.” *Youmatter*, youmatter.world, 11 July 2017, <https://youmatter.world/fr/Smart-city-ville-durable-innovation/>. Dernier accès le 3 June 2022.
- [38] G. H. Brundtland, *Our common future: The world commission on environment and development*, Oxford University Press, Oxford, 1987.
- [39] “TLFi.” *TLFi*, atilf.atilf.fr, <http://atilf.atilf.fr/>. Dernier accès le 3 June 2022.
- [40] Kummitha, R. K. R., & Crutzen, N. (2017). How do we understand Smart cities? An evolutionary perspective. *Cities*, 67, 43-52.
- [41] Tippmann, V., & Von Radecki, A. (2016). La vision «Morgenstadt». La ville de demain : des partenariats d’innovation pour un développement urbain durable. In *Annales des Mines-Responsabilité et environnement* (No. 4, pp. 74-78). FFE.
- [42] Berte, D. R. (2018, May). Defining the IoT. In *Proceedings of the international conference on business excellence* (Vol. 12, No. 1, pp. 118-128).
- [43] García-Peñalvo, F. J., De Figuerola, C. G., & Merlo, J. A. (2010). Open knowledge: Challenges and facts. *Online information review*.
- [44] Vanolo, A. (2014). Smartmentality: The Smart city as disciplinary strategy. *Urban studies*, 51(5), 883-898.
- [45] Etzkowitz, H., & Zhou, C. (2006). Triple Helix twins: innovation and sustainability. *Science and public policy*, 33(1), 77-83.
- [46] Giffinger, R., Fertner, C., Kramar, H., & Meijers, E. (2007). City-ranking of European medium-sized cities. *Cent. Reg. Sci. Vienna UT*, 9, 1-12.
- [47] Kumar, V., & Dahiya, B. (2017). Smart economy in Smart cities. In *Smart economy in Smart cities* (pp. 3-76). Springer, Singapore.
- [48] Coulée, C., Dieuzeide, M., & Bréchet, T. Smart cities: quelles sont les principales menaces?. *CHAPITRE, 1*, 2017-2018.
- [49] Anthopoulos, L., Janssen, M., & Weerakkody, V. (2019). A Unified Smart city Model (USCM) for Smart city conceptualization and benchmarking. *Smart cities and Smart spaces: Concepts, methodologies, tools, and applications*, 247-264.
- [50] BENBLIDIA, M. A., ESSEGHIR, M., & BRIK, B. (2020). Gestion intelligente des ressources dans un système Smart Grid-Cloud pour une meilleure efficacité énergétique. *Gestion et contrôle intelligents des réseaux: Sécurité intelligente, optimisation multicritères, Cloud Computing, Internet of Vehicles, radio intelligente*, 181.
- [51] Buckman, A. H., Mayfield, M., & Beck, S. B. (2014). What is a Smart building?. *Smart and Sustainable Built Environment*.

- [52] Dustdar, S., Nastić, S., & Šćekić, O. (2017). Smart cities. In *The Internet of Things, People and Systems*. Berlin/Heidelberg, Germany: Springer.
- [53] Desdemoustier, J., & Crutzen, N. (2015). *Smart Cities en Belgique: Analyse qualitative de 11 projets*. Smart city Institute.
- [54] Dieuzeide, M., Coulée, C., & Bréchet, T. Smart cities: quelles sont les principales menaces?.
- [55] Abbas, Rasoul Adnan. "Figure 1-1 from A COMPARISON OF SMART CITY INDICATORS FOR THREE TOP TEN US CITIES | Semantic Scholar." *Figure 1-1 from A COMPARISON OF SMART CITY INDICATORS FOR THREE TOP TEN US CITIES | Semantic Scholar*, www.semanticscholar.org, 1 Jan. 2019, <https://www.semanticscholar.org/paper/A-COMPARISON-OF-SMART-CITY-INDICATORS-FOR-THREE-TOP-Abbas/b0d2d1e96e908ce04a1c118e906eef92ee7e11ba/figure/0>. Dernier accès le 3 June 2022.
- [56] Lim, C., Kim, K. J., & Maglio, P. P. (2018). Smart cities with Big Data: Reference models, challenges, and considerations. *Cities*, 82, 86-99.
- [57] Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of Big Data to Smart cities. *Journal of Internet Services and Applications*, 6(1), 1-15.
- [58] Clarke, A., & Steele, R. (2011, December). How personal fitness data can be re-used by Smart cities. In *2011 Seventh International Conference on Intelligent Sensors, Sensor Networks and Information Processing* (pp. 395-400). IEEE.
- [59] Bibri, S. E. (2018). A foundational framework for Smart sustainable city development: Theoretical, disciplinary, and discursive dimensions and their synergies. *Sustainable Cities and Society*, 38, 758-794.
- [60] Afzalan, N., Sanchez, T. W., & Evans-Cowley, J. (2017). Creating Smarter cities: Considerations for selecting online participatory tools. *Cities*, 67, 21-30.
- [61] Aina, Y. A. (2017). Achieving Smart sustainable cities with GeoICT support: The Saudi evolving Smart cities. *Cities*, 71, 49-58.
- [62] Cicirelli, F., Guerrieri, A., Spezzano, G., & Vinci, A. (2017). An edge-based platform for dynamic Smart city applications. *Future Generation Computer Systems*, 76, 106-118.
- [63] George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493-1507.
- [64] Barns, S. (2016). Mine your data: Open data, digital strategies and entrepreneurial governance by code. *Urban geography*, 37(4), 554-571.

- [65] Li, Z., & Shahidehpour, M. (2017). Deployment of cybersecurity for managing traffic efficiency and safety in Smart cities. *The Electricity Journal*, 30(4), 52-61.
- [66] KLID (2014). Study on developing strategies for Big Data application in private and public sector through case-study analysis. Korea Local Information Research & Development Institute.
- [67] Kitchin, R. (2014). The real-time city? Big Data and Smart urbanism. *GeoJournal*, 79(1), 1-14.
- [68] Apulu, I., & Latham, A. (2011). An evaluation of the impact of Information and Communication Technologies: Two case study examples. *International Business Research*, 4(3), 3-9.
- [69] Bloom, N., Garicano, L., Sadun, R., & Van Reenen, J. (2014). The distinct effects of information technology and communication technology on firm organization. *Management Science*, 60(12), 2859-2885.
- [70] Mikre, F. (2011). The roles of information communication technologies in education: Review article with emphasis to the computer and internet. *Ethiopian Journal of Education and Sciences*, 6(2), 109-126.
- [71] Van, D. P., Fujiwara, T., Tho, B. L., Toan, P. P. S., & Minh, G. H. (2020). A review of anaerobic digestion systems for biodegradable waste: Configurations, operating parameters, and current trends. *Environmental Engineering Research*, 25(1), 1-17.
- [72] Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of Big Data to Smart cities. *Journal of Internet Services and Applications*, 6(1), 1-15.
- [73] KAROUANI, Y., & ZIYATI, H. Big Data dans l'architecture de l'IoT pour les villes intelligentes.
- [74] Wei, Y., & Blake, M. B. (2010). Service-oriented computing and cloud computing: Challenges and opportunities. *IEEE Internet Computing*, 14(6), 72-75.
- [75] N. Benhamida, L. Bouallouche-Medjkoune, D. Aïssani *et al.*(2022). Dynamic Architecture for Collaborative Distributed Storage of Collected Data in Fog Environments. *Wireless Personal Communication* (4) **123**, 3511–3537.
- [76] Steffemel, L. A. (2018). Improving the Performance of Fog Computing Through the Use of Data Locality. 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD), Lyon, France, pp. 217-224, Doi: 10.1109/CAHPC.2018.8645879.

- [77]Hu, P., Dhelim, S., Ning, H., & Qiu, T. (2017). Survey on fog computing: architecture, key technologies, applications and open issues. *Journal of Network and Computer Applications*, 98, pp. 27–42.
- [78]Sharma, P.K., Chen, M.Y., & Park, J.H. (2018). A software defined fog node based distributed blockchain cloud architecture for IoT. *IEEE Access*, 6, pp. 115–124, Doi: 10.1109/ACCESS.2017.2757955.
- [79] Kim, D., Son, J., Seo, D., Kim, Y., Kim, H., & Seo, J. T. (2020). A novel transparent and auditable Fog-assisted cloud storage with compensation mechanism. *Tsinghua Science and Technology*, 25(1), pp. 28-43, Doi: 10.26599/TST.2019.9010025.
- [80] BEN MOULA, A. (2022). 7. Les différences entre Cloud Computing, Edge Computing et Fog Computing.
- [81] “CupCarbon IoT Simulator.” *CupCarbon IoT Simulator*, CupCarbon.com, <http://CupCarbon.com/>. Dernier accès le 3 June 2022.