

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/340085425>

# Big Data et ville intelligente

Article · March 2020

CITATIONS

0

READS

825

1 author:



Niama Elkhbir

École Nationale Supérieure de l'Electronique et de ses Applications

3 PUBLICATIONS 0 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Logiciel de modélisation des systèmes satellitaires [View project](#)



# Rapport Big Data et ville intelligente

Réalisé par:

Niama ELKHBIR

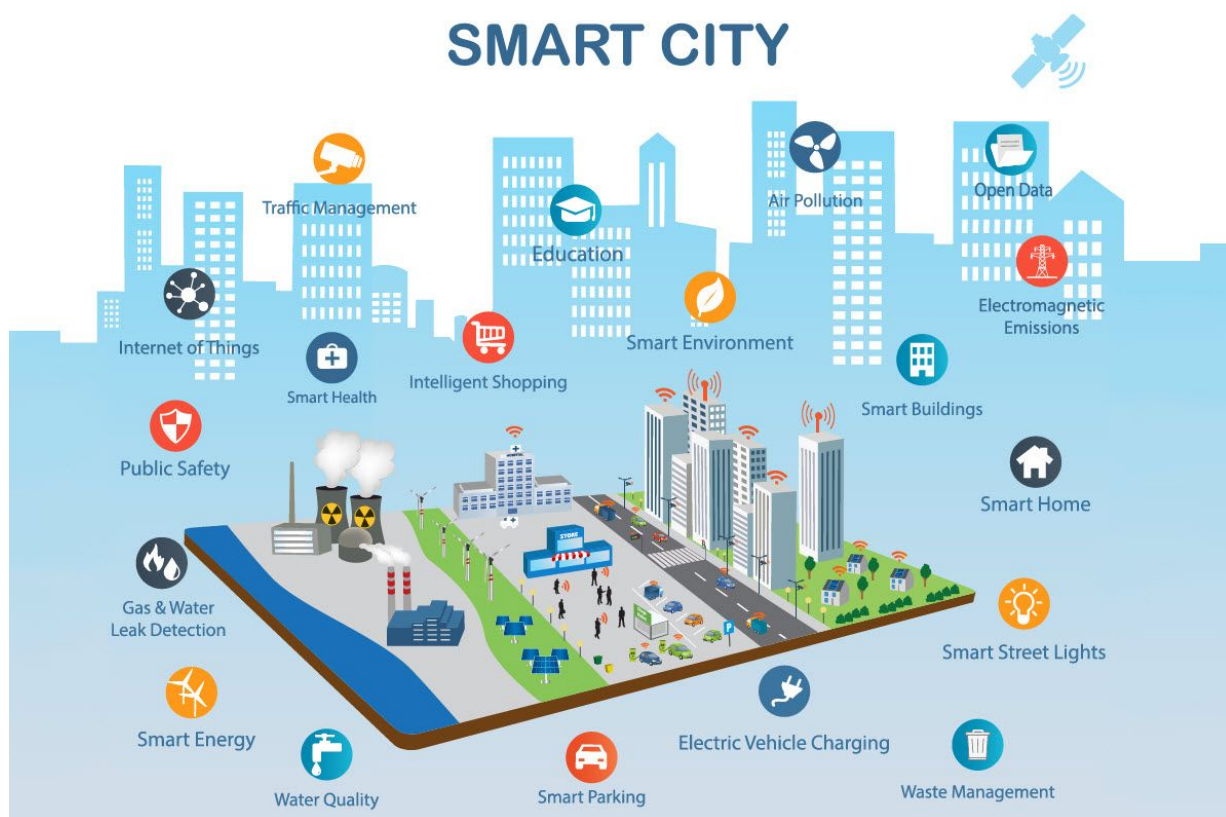
Assem AMEZIANE

# **Sommaire**

- I. Présentation
- II. Big Data dans la ville intelligente
- III. Méthodes utilisées pour traiter le Big Data
- IV. Infrastructure pour le traitement du Big Data
- V. Conclusion
- VI. Références

## **I. Présentation**

Les villes intelligentes sont devenues sujet de discussion et de recherche dans plusieurs universités, entreprises et même gouvernements, ce qui a attiré notre curiosité et nous a amené à faire cette petite recherche. Les villes numériques désignent des villes qui intègrent les technologies de l'information et de la communication (TIC) et divers dispositifs physiques connectés au réseau (l'Internet des objets ou IoT) pour optimiser l'efficacité des opérations et des services urbains et se connecter aux citoyens [1]. Jusqu'à présent, il n'existe pas de définition universelle ou standard de la ville intelligente mais toutes les définitions existantes affirment que la ville intelligente est une solution de vie intégrée qui relie de nombreux aspects de la vie tels que l'énergie, le transport et les bâtiments de manière intelligente et efficace afin d'améliorer la qualité de vie des citoyens de cette ville.



*figure0: La ville intelligente et ses interactions*

Actuellement, ce qui préoccupe le plus les gouvernements du monde entier intéressés par le concept de la ville intelligente est le coût d'acquisition de cette ville, en raison de la diversité des capacités financières et de la rareté des ressources, naturelles ou humaines. La disponibilité, la taille et les capacités de ces ressources constituent l'un des défis de la

construction et du maintien d'une ville intelligente, mais les systèmes de réglementation peuvent également affecter considérablement les chances de réussite. Il faut aussi souligner que les défis techniques exigent une grande capacité de stockage vu que les données sont générées à partir de multiples sources. Ces défis amènent à penser à exploiter les Big Data.

Les données sont aujourd'hui existantes partout; sur les téléphones intelligents, les ordinateurs, les capteurs environnementaux, les appareils photo, les GPS et même sur les gens. Ces dernières années, diverses applications telles que les sites de médias sociaux, les photos et vidéos numériques, les transactions commerciales, les applications publicitaires, les jeux et bien d'autres ont contribué à accélérer la production de données [2].

Parmi les cas d'usage du Big Data dans la ville intelligente, on cite [3]:

1. Le transport : lors d'événements publics de grande audience, les opérateurs ferroviaires se trouvent en face à plusieurs millions de données qui peuvent être gérées grâce aux données et technologies analytiques pour une circulation fluide et une organisation efficace. Un exemple concret d'application a été fait pendant les Jeux olympiques de Londres 2012, où le réseau de transport public a dû gérer 18 millions de trajets effectués par les spectateurs venus de toute la ville.
2. La planification urbaine : le Big Data permet d'augmenter l'efficacité lors de la planification et la fabrication d'immeubles et l'aménagement d'espaces urbains. Les données permettent de cartographier et de prédire l'impact d'une infrastructure sur l'espace urbain avec une précision élevée. Il est également possible de développer des modèles pour maximiser l'accès à certaines zones ou services tout en minimisant les risques de surcharge d'infrastructure.
3. La sécurité : Les analyses prédictives ont été utilisées dans plusieurs villes autour du monde pour aider à prédire à quel endroit les crimes risquent de survenir, en se basant sur les données historiques et géographiques. Dans des villes comme Londres, Los Angeles, et Chicago, cette initiative a porté ses fruits.
4. Dépenses efficaces : Comme nous l'avons cité précédemment, le problème principal des villes intelligentes est que de grandes quantités d'argent sont dépensées pour les travaux. Les petits changements ou le remodelage des paysages peuvent être considérés comme un gaspillage de l'argent public. En utilisant les technologies du Big Data, il est possible de vérifier à l'avance l'impact des dépenses. Ainsi, il est possible de déterminer comment dépenser l'argent efficacement pour améliorer la ville à moindre coût.

## **II. Big Data et la ville intelligente**

### **1. Types de données**

Dans les villes intelligentes, le flux des données est grand, leurs sources sont différentes et leurs types aussi [4], ainsi on distingue les données traditionnelles et structurées des systèmes d'entreprise, comme par exemple les prévisions météorologiques, les données démographiques du gouvernement et aussi les statistiques sur les performances des transports publics. On cite aussi les données provenant de toutes sortes de médias sociaux. Et finalement, on n'oublie pas les données Machine-to-Machine (M2M).

### **2. Diversité des données**

Les villes intelligentes génèrent une variété de données non structurées (selon la nature de la source de données). On distingue [5] les données chronologiques (bourse, voitures intelligentes, maisons intelligentes), les données en continu (capteur, systèmes de surveillance, trafic Internet), les données séquentielles (systèmes de vente au détail, ADN humain, structure des protéines), les données graphiques (réseaux sociaux, sites web, systèmes de santé), les données spatiales (données SIG, imagerie médicale, données satellites) et aussi les données multimédia (systèmes de surveillance, maisons intelligentes).

### **3. Open Data**

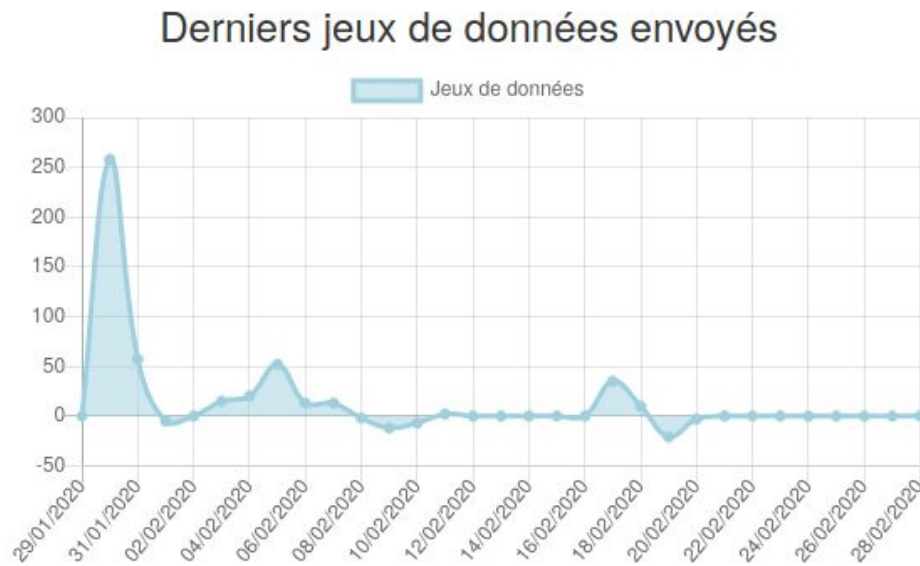
Plusieurs initiatives sont prises par la métropole, par des entreprises ou des individus consistant à produire des cartes pour mettre les données en scène et de façon intelligible. D'après Benoît Chabrier [6], délégué régional au numérique à la préfecture d'Occitanie, les données ouvertes sont très peu exploitées. Les associations et la population manquent encore cruellement de culture du numérique en général et de la donnée en particulier. Plusieurs projets ont été mis en place pour rendre la donnée publique brute intelligible par tous comme par exemple le projet Dataclac, mené en Occitanie, un des onze projets Investissements d'avenir sur l'open data des territoires. D'autre part, il existe des sites officiels d'Open Data comme le site du gouvernement français [data.gouv.fr](http://data.gouv.fr), ou le site [www.data.gov](http://www.data.gov), et plusieurs sites internet proposent des solutions de partage des données, comme par exemple OpenDataSoft...

### **4. Exemple de statistiques de données**

D'après [7] on peut visualiser les nombres d'utilisateurs, de jeux de données, de ressources (figure1), et aussi suivre en temps réel leur évolution(figure2).



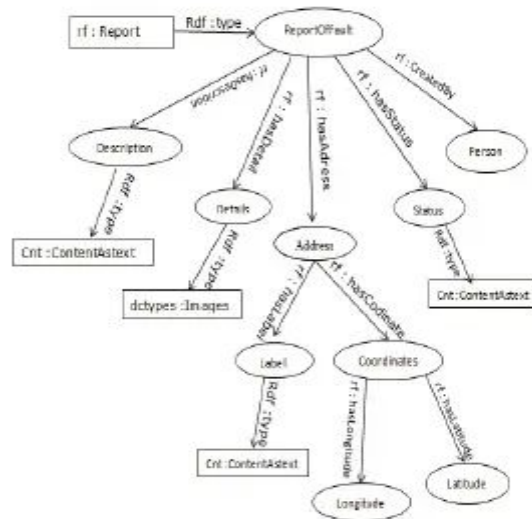
*figure1: statistiques officielles depuis le site data.gouv.fr*



*figure2: derniers jeux de données envoyés depuis le site data.gouv.fr*

## **5. Linked Data**

D'après [8], une approche a été faite en définissant une ontologie OWL qui décrit un modèle générique pour une entité. La classe principale est appelée ReportOfFault, qui contient la personne qui peut détecter et informer l'administrateur. La classe Détail représente les images et les photos de défaut, et la classe Description inclut le commentaire fait par toute personne. La classe Status est introduite pour déterminer le statut de chaque défaut (commencé, en réparation ou terminé). Enfin, la classe Adresse est définie pour représenter l'adresse du défaut, cette classe comporte deux sous-classes à savoir, Étiquette (adresse du défaut) et Coordonnées (Latitude et longitude).



*figure3: Architecture de l'ontologie*

## **6. Big Data et vie privée**

Avec tous ces avantages que présente la ville intelligente, il y'a sûrement un prix à payer. Il nous vient tout de suite à l'esprit la fameuse question: Risquons nous notre vie privée ? Serions-nous exposés ou même contrôlés?

Nous sommes identifiées par différents moyens [11]: informations biographiques, données biométriques (visage, empreintes digitales, scanners, Iris...), données comportementales, données de voyage et informations bancaires. Les gouvernements ont toujours été obsédés par la collecte de données sur les citoyens, par exemple l'agence de sécurité de l'État est-allemand connue sous le nom de Stasi jusqu'à la chute du mur de Berlin en 1989 a mis en place un système de surveillance et d'espionnage de ses citoyens, ils ont utilisé à la fois des instruments techniques et l'intelligence humaine, les téléphones ont été mis sur écoute, des machines spéciales pour ouvrir les lettres, prendre des photos des gens et même leur odeur ! Les espions et les agents rapportaient non seulement ce que les gens faisaient, mais aussi ce qu'ils pensaient et planifient. À l'époque, le web et les smartphones n'étaient pas encore inventés, mais aujourd'hui, la technologie facilite encore plus la collecte, le stockage et, surtout, l'analyse des données.

Ce que nous ne savons pas, c'est que les énormes quantités de données collectées sur les utilisateurs présentent une mine d'or, facilitant le contrôle des gens et les rendant vulnérables. Parmi les manières d'exploiter ces données, on cite:

1. L'e-commerce intelligent : A travers les systèmes de recommandation, le ciblage comportemental, le profilage de la clientèle, le marketing ciblé et bien d'autres sont une série de technologies utilisées "contre" les utilisateurs à des fins commerciales. Ce



sont des techniques inventées pour enquêter sur les préférences, les tendances, la personnalité, la religion, l'orientation sexuelle et les opinions politiques des individus à l'insu des utilisateurs et tirer parti de ces informations pour améliorer le marketing et le secteur de la publicité. Il n'y a rien de mal à apporter une meilleure expérience à l'utilisateur et à aider les gens à trouver facilement ce qu'ils veulent, mais le mauvais côté est lorsque ces informations sont utilisées pour contrôler la liberté de choix des utilisateurs, les forcer à acheter ce dont ils n'ont pas besoin, les transformer en obsédés de la consommation et les hypnotiser à payer le plus possible afin d'augmenter les ventes et les revenus des entreprises.

2. Les moteurs de recherche intelligents : Les moteurs de recherche comme Google, par exemple, limitent nos résultats de recherche et les personnalisent. Les résultats de recherche sur un même sujet de recherche sont différents et dépendent bien sûr des informations collectées sur chaque utilisateur. Comme les algorithmes de filtrage qu'ils utilisent décident de ce qu'il faut montrer et de ce qu'il faut nous cacher, la capacité de voir d'autres points de vue est écartée et, par conséquent, nous sommes privés de cet équilibre naturel des choses non seulement pertinentes mais aussi ils nous rendent loins de ce qui se passe réellement dans le monde réel.
3. La politique intelligente : La signification fondamentale des élections politiques est le droit du citoyen à choisir son candidat préféré en fonction du flux d'informations qu'il reçoit. Ainsi, lorsque ce citoyen utilise des services en ligne, par exemple les médias sociaux, les informations, les moteurs de recherche, qui l'obligent à ne voir qu'une seule facette de l'histoire, les élections semblent devenir une grosse plaisanterie dans notre vie.
4. La transportation intelligente : Basée sur l'utilisation d'appareils d'enregistrement pour collecter des données, notamment sur la vitesse, l'accélération, le freinage, le port de la ceinture de sécurité, l'état du véhicule, le déploiement des airbags, numéros de téléphone et de contact, messages, GPS, données de voyage, site d'origine... Les systèmes de transport équipés d'instruments offrent des cibles appropriées pour un délinquant motivé par la traque ou la violence domestique. La vie privée de la victime/cible est fortement compromise dans la mesure où l'accès aux systèmes de transport fournit au délinquant des informations presque complètes sur le lieu, le moment et la durée de la visite de la victime/cible à un endroit particulier. Il peut fournir des informations supplémentaires sur les personnes que la victime/cible a appelées. Cette atteinte à la vie privée constitue un risque majeur pour la sécurité. Une fois que le délinquant motivé a un profil et un emplacement sur la victime/cible à tout moment, il sait à quel moment cette victime/cible serait la plus vulnérable à une attaque physique.

### **III. Méthodes utilisées pour traiter le Big Data**

Le volume massif des données générées par les villes intelligentes nécessite qu'elles soient collectées, gérées et analysées pour obtenir des informations utiles et fournir les fonctionnalités requises. À cet égard, l'analyse des méga données joue un rôle important pour extraire des informations utiles et pour faire des prédictions, des identifications des tendances ou prendre des décisions.

Les méthodes utilisées pour traiter ce volume massif de données des villes intelligentes sont :

#### **1. Data visualisation**

C'est une méthode clé du Big Data car elle permet aux acteurs humains de comprendre l'importance des données en les mettant dans un contexte visuel. Les techniques de visualisation abstraite fournissent un résumé des données à grande échelle avant de les rendre aux unités de visualisation. Les exemples sont les cubes de données, le regroupement d'histogrammes, l'agrégation hiérarchique, etc. La visualisation interactive englobe des techniques qui permettent des visualisations et l'interaction des utilisateurs en temps réel.

#### **2. Data analytics**

Une grande variété d'algorithmes d'apprentissage automatique sont utilisés pour extraire des informations bien informées des mégadonnées générées par les villes intelligentes pour faire des prédictions, identifier les tendances, découvrir des informations cachées ou prendre des décisions[9]. Selon les exigences, une approche analytique appropriée doit être choisie. Des algorithmes supervisés sont utilisés à des fins de classification et de prédiction / régression, tandis que des algorithmes non supervisés sont généralement utilisés pour le regroupement et la séparation.

Les algorithmes semi-supervisés sont utilisés sur des ensembles de données non étiquetés sur lesquels les algorithmes supervisés traditionnels ne peuvent pas être appliqués. Les techniques d'apprentissage semi-supervisées utilisent la similitude structurelle entre les données étiquetées et non étiquetées de manière efficace pour généraliser la cartographie fonctionnelle sur de grands ensembles de données. Les méthodes d'apprentissage par renforcement tentent de créer des fonctions de cartographie appropriées entre les observations et les actions dans le but de maximiser une fonction de récompense.

### **3. Data variety**

Les villes intelligentes génèrent une variété de données non structurées (selon la nature de la source de données). Les données de séries chronologiques sont des séquences de valeurs ou d'événements obtenus à partir de mesures répétées dans le temps. Les données de streaming se réfèrent à celles qui arrivent en permanence, par exemple les données des capteurs, le trafic Internet, etc.

Les données de séquence sont constituées d'éléments ordonnés ou d'événements qui sont enregistrés avec ou sans une notion concrète du temps [10]. Les informations qui proviennent des réseaux sociaux, du World Wide Web, des réseaux du corps humain, etc. sont naturellement adaptées pour être modélisées sous la forme d'une structure de données graphique, et sont donc appelés les données du graphique. Les données spatiales font référence aux informations obtenues à partir de sources telles que la télédétection, les systèmes d'information géographique ou les données d'imagerie médicale. Enfin, les données multimédias comprennent des images, des vidéos et du son. Chacun des types de données mentionnés ici a ses propres caractéristiques uniques et est analysé à l'aide de différentes techniques d'exploration de données.

## **IV. Infrastructure pour le traitement du Big Data**

### **1. Les infrastructures informatique**

Cela fait référence aux différentes plateformes de traitement qui sont normalement utilisées pour les grands ensembles de données provenant des villes intelligentes. En fonction des besoins des données, les infrastructures informatiques peuvent traiter les données en temps réel, quasi réel ou en mode batch. Par exemple, Hadoop est couramment utilisé pour le traitement par lots, tandis que Spark est utilisé pour le traitement en temps réel.

### **2. Les infrastructures de stockage**

Les données collectées dans les villes intelligentes sont très variées, allant du multimédia au texte. Une grande partie des données des capteurs ne sont pas structurées par nature, et donc en plus de la structure de base de données relationnelle normale, d'autres types de bases de données sont nécessaires. Par conséquent, l'infrastructure de stockage est choisie pour déterminer le type de stockage nécessaire en fonction du type de Big Data.

Des systèmes de stockage sont basés sur SQL comme Oracle, MySQL, etc., d'autres sont basés sur NoSQL (MongoDB, Aerospike, HBase, Cassandra, etc.) et NewSQL (HStore, VoltDB, etc.).

## **V. Conclusion**

Durant ce travail, on a présenté la ville intelligente orientée Big Data. Nous avons montré l'utilité du Big Data pour développer des villes intelligentes basées sur l'infrastructure informatique, l'infrastructure de stockage, la variété des données, l'analyse des données et la visualisation des données pour la compréhension des lecteurs. De plus, nous avons vu les principales plateformes d'analyse du Big Data tel que Hadoop, spark .

En outre, aujourd'hui des cas ont été sélectionnés sur les villes intelligentes à travers le monde pour révéler une tendance à la hausse des déploiements de villes intelligentes. En fin de compte, plusieurs défis de recherche ouverts ont été discutés tels que la sécurité, confidentialité, l'intégration des données et l'analyse des données, qui exigent l'attention de la communauté des chercheurs et devraient ouvrir la voie à de futurs travaux.

## **VI. Références**

- [1] [https://fr.wikipedia.org/wiki/Ville\\_intelligente](https://fr.wikipedia.org/wiki/Ville_intelligente).
- [2] han Z, Anjum A, Kiani SL. Cloud Based Big Data Analytics for Smart Future Cities. In Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. IEEE Computer Society.
- [3] <https://www.lebigdata.fr/smart-cities-big-data>.
- [4] <https://www.information-age.com/three-types-data-will-need-work-together-run-smart-city-123460068/>.
- [5] Big Data in Smart-Cities: Current Research and Challenges. Available from: [https://www.researchgate.net/publication/329916984\\_Big\\_Data\\_in\\_Smart-Cities\\_Current\\_Research\\_and\\_Challenges](https://www.researchgate.net/publication/329916984_Big_Data_in_Smart-Cities_Current_Research_and_Challenges) [accessed Feb 28 2020].
- [6] <https://dataanalyticspost.com/de-lopen-data-city-a-smart-city/>.
- [7] [data.gouv.fr](http://data.gouv.fr).
- [8] Smart Cities Based on Web Semantic Technologies [https://www.academia.edu/34992357/Smart\\_Cities\\_Based\\_on\\_Web\\_Semantic\\_Technologies](https://www.academia.edu/34992357/Smart_Cities_Based_on_Web_Semantic_Technologies).
- [9] Mohammadi M, et al., "Enabling Cognitive Smart cities using Big Data and Machine Learning: Approaches and Challenges," IEEE Communications Magazine, vol. 56(2), pp. 94-101, 2018.
- [10] Mahdavinejad MS et al., "Machine Learning for Internet of Things Data Analysis: A Survey," Digital Communication and Networks, vol. 4(3), pp. 161-175, 2018.
- [11] Age of Big Data and Smart Cities: Privacy Trade-Off. Available from: [https://www.researchgate.net/publication/267759778\\_Age\\_of\\_Big\\_Data\\_and\\_Smart\\_Cities\\_Privacy\\_Trade-Off](https://www.researchgate.net/publication/267759778_Age_of_Big_Data_and_Smart_Cities_Privacy_Trade-Off).