

# Planning of smart cities: Performance improvement using big data analytics approach

Bhagya Nathali Silva, Murad Khan, Changsu Jung, Jihun Seo, Yongtak Yoon, Jinbae Kim, Seungpyo Jin, Jeongyun Kang, Kijun Han\*

**Abstract**— The concept of smart city is widely favored, as it enhances the quality of urban citizens' life, involving multiple disciplines. Consequent to the complex urban networks, data processing complexity has increased significantly. Thus, it creates a crucial demand to facilitate autonomous decision-making and real-time data processing and analysis of smart cities. Therefore, in this paper we propose a smart city framework based on Big Data analytics. The proposed framework operates in three levels 1) Data generation and acquisition level, 2) Data management and processing level, and 3) Application level. Moreover, we analyzed the water consumption, traffic congestion, and air pollution data of Surrey (Canada) and Aarhus (Denmark) cities to determine the threshold values for data filtering. The analysis shows that the proposed architecture offers useful insights to the community development authorities to improve the existing smart city architecture.

**Keywords**— Smart city, big data, urban planning.

## I. Introduction

The revolutionized growth of smart devices has encouraged the concept of “connected everyday objects” over the existing network. The growth spike of the connected devices has led towards the boundless expansion of the conventional networks, inspiring the renaissance of the web as the third wave “Internet of Things (IoT)”. The IoT is an unceasingly growing network of uniquely addressable heterogeneous devices. Substantially, it identifies and shares data autonomously or with minimal human interaction. The dramatic increase in the number of embedded devices has played a vital role to draw the attention of multiple interest groups towards IoT. Consequent to the extensive attention, IoT has become a mature concept, pioneering the novel applications i.e. smart home, smart city, smart grid, smart health, etc. The smart city notion is initially coined with the aim of utilizing public resources and services efficiently, to increase the quality of services offered to the urban citizens [1]. In fact, the smart cities optimize the conventional public services i.e. transportation, parking, surveillance, health care, electricity, etc. with aid of the autonomous data collection facilitated through heterogeneous devices connected to the urban IoT. In order to serve real-time service requests, timely processing of the colossal amount of data is essential. challenges remarkably enhances the quality of life of the smart

city citizens. However, general data processing and analytical mechanisms are unable to satisfy the real-time processing as the volume of data increase immensely. Hence, the ideal first step towards a smarter city is to collaborate with Big Data analytics, to assure, flexible and real-time intelligent decision-making [2]. Consequently, the novel approach of addressing the existing.

The rapid development of the smart city system diverts the focus of the many researchers and architects towards an efficient communication and standard architectural design. Similarly, various schemes have been proposed following a thorough experimentation and test bed based simulations to overcome the challenges present in the current literature. A scheme based on experimenting a complete set of smart city services on various test bed modules have been proposed in [3]. The author developed the physical implementation of a large-scale IoT infrastructure in Santander city. The experimental facility is designed much user-friendly so that the experimenter can test the facility in different urban environments, smart city planning, etc. A variety of new mechanisms was developed following the Santander city requirements. These mechanisms include mobility support, security and surveillance systems, large-scale support, scalability heterogeneity, etc. in a smart city environment. The testbed results show that the proposed architecture covers several challenges in the current literature. However, the data collected from various sensors is not tested for future urban planning and designing. Therefore, the architecture can guarantee better services in one environment but may be poor performance in another environment. Similarly, the demands of the user in an IoT-based smart environment rapidly changes. Hence, it decreases the chances of understanding the context and dynamicity of the IoT smart user. On the other hand, the IoT is not yet matured to deploy it as generic standard for designing smart services such as smart homes, cities, etc. because of the two main reasons, 1) The current IoT-based solutions are limited to specific application domain and 2) new technologies and optimization techniques are good in one area but maybe not in another. For example, Wireless Sensor Networks (WSN) suffered from high packet loss in a heterogeneous wireless environment. Similarly, Wireless Local Area Network (WLAN) can provide low-cost services but provides a narrow coverage compared to other technologies. Therefore, the researchers come up with several solutions, which ultimately leads to a generic communication model and covering a wide set of services [4] [5]. In addition, a generic communication model can be achieved by integrating the WSN with the existing infrastructure. Thus, this will help in achieving a real internet of things [6].

Bhagya Nathali Silva, Murad Khan, Changsu Jung, Jihun Seo, Yongtak Yoon, Jinbae Kim, Seungpyo Jin, Jeongyun Kang, Kijun Han\*  
School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea

In this paper, we propose a realistic smart city architecture integrating Big Data analytics. The proposed scheme facilitates real-time intelligent and autonomous decisions, fast and reliable data collection, smart grid, and user-centric customization of energy consumption. Thus, the spotlight of the study is focused on accurate and prompt decision-making. Henceforth, the enormous amount of heterogeneous data is organized via the data fusion techniques to speed up the data processing in Big Data analytics. In this work, we use Hadoop as the storage and processing medium for the collected Big Data. Hadoop processes the organized data to make timely decisions. Sequentially, the decisions are followed by respective events or actions to be performed. The rest of the paper is organized as follows. Section 2, give a brief description of the proposed architecture. The results and analysis section is presented in Section 3. Finally, the conclusion is outlined in section 4.

## II. Proposed Scheme

The proposed smart city architecture comprises with three levels i.e. 1) Data generation and acquisition level, 2) Data management and processing level, and 3) Application level. The data collection level is the bottom level and liable for the collection and acquisition of heterogeneous data. Meanwhile, the middle layer processes heterogeneous data to elicit intelligent decisions autonomously and redirect towards the top level. The application level resides on the top and initiates action execution in accordance with the decisions and events received from the middle layer.

### A. Data Generation and Acquisition Level

Smart city involves a prodigious amount of data, complex and comprehensive computations, and multiple application domains. The concept of the smart city is to optimize residential resources, reduce traffic congestion, efficient healthcare services, water management, etc. Data acquisition related to daily activities become vital, to achieve the aforementioned goals. Indeed, the acquisition of data is challenging due to the large amount of data created by people and other connected objects. Therefore, effective mechanisms are employed to sense and identify a phenomenon of interest from the real world and converts into digital form for further processing. Low cost and energy efficient sensors are accepted widely to acquire heterogeneous data from the urban IoT. As the number of sensors connected within the city increases, the city becomes smarter [7]. Thus, the deployment of large number of disparate sensors in the city becomes the initial step of the proposed scheme. These sensors collect real-time data from the instated environment. The type of collected data differs from the deployed context i.e. smart home, vehicular transportation system, healthcare management system, and meteorology and water management system.

The prime objective of the smart home is towards the energy management of residential buildings. The sensors attached to the home appliances transmit real-time energy consumption of a particular household to the middle layer. As a result of data processing, the user becomes aware about the current energy consumption via the smart community development. The utmost goal of the vehicular transportation system is to reduce the city traffic congestion. The sensors implanted on roadways collect vehicle entrance and departure of the particular road, to determine the level of road congestion. Moreover, the meteorology department determines weather and other environmental parameters. For example, the sensors implanted within the city suburbs sense carbon monoxide concentration of a certain location and communicate to the middle level.

In order to transfer sensed data to the data management and processing level, multiple communication technologies are used i.e. ZigBee, Bluetooth, Wi-Fi, and data and cellular networks.

### B. Data Management and Processing Level

The data management and processing level bridges the data acquisition with the application level. This layer is the brain of the proposed framework, as it filters valuable data, analyzes, processes, stores, makes context aware decisions, and finally generates corresponding events. Multiple modalities are embedded into this layer to perform aforementioned tasks. The framework facilitates offline and real-time data fusion to filter the valuable data from the enormous amount of the sensed data. Data analysis utilizes MapReduce paradigm, while data manipulation and storing occupies hadoop distributed file system (HDFS), HBASE, and HIVE.

The data processing efficiency of the smart city is enhanced with the data filtration facilitated via fusion techniques. The proposed scheme occupies the Kalman filter as the fusion model. In general, the Kalman filter removes noise from the sensed data. Thence, the smart city framework distinguishes unnecessary bulk of data as the noise of data. Henceforth, the Kalman filter is manipulated to determine valuable data, by assigning threshold values for each parameter of sensed data. Therefore, the Kalman Filter determines the congested roads for further analysis. Ultimately, the analysis speed is increased due to the reduction of futile data.

Consequently, the fused data goes under analysis. The proposed scheme employs Hadoop as the data storing and processing framework. Therefore, proposed scheme incorporates MapReduce for the data analysis. The input data received from data fusion is converted into another set of data in the mapping process of MapReduce. Thereupon, the reduce task combines data from the mapping into a set of values reduced in amount.

Data storing and processing is another vital task of the middle level. Multiple techniques facilitate this demand of the

proposed framework i.e. HDFS, HBASE, HIVE, etc. as shown in Figure 1. The HDFS facilitates the distributed storage requirement of the smart city. Moreover, HDFS facilitates the scalability demand of Big Data processing. The proposed scheme relies on the real-time read/write facility over data throughout the cluster, in order to facilitate autonomous intelligent decision making. Thus, HBASE is used on top of the HDFS, due to its higher fault tolerance, usability, and faster processing. HIVE provides the querying and managing facility over the large amount of data resides on the Hadoop cluster. Since SQL cannot be used to query on HIVE, we have used HiveQL to query the data on Hadoop cluster. Finally, the derived intelligent decisions generates corresponding events. Simultaneously, the generated events and decisions are transferred to the application level of the smart city framework.

### C. Application Level

Application level is the top layer of the framework, which ensures the execution of actions corresponding to the autonomous intelligent decisions. As the data processing level transfers an event, the target application in the application level becomes liable for that particular event execution. In order, each application on the top level is implemented to handle numerous events that ensure readily on demand operation of the smart city. Assume that the sensors implanted on a particular city observe a street congestion. The congestion level is analyzed at the data processing level. Subsequently, the data processing level generates alternative paths to avoid congestion. At the same time, the alternative paths are communicated to the smart traffic control department in the application level. The application conveys the alternative paths to the travelers accordingly. Moreover, the paths are notified to the potential travelers, whom may enter into the congested street. The smart traffic control department determines the fact by the GPS destination check and the current positioning of the vehicle.

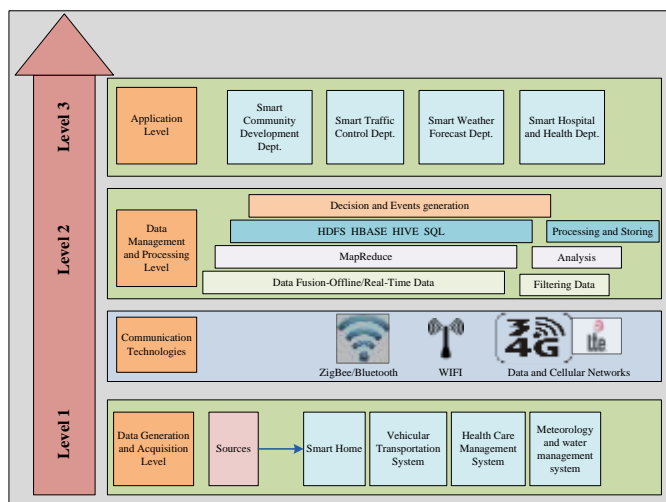


Figure 1 Working of the proposed architecture

## III. Results and Data Analysis

In order to design an efficient smart city architecture, the data obtain from various data sources is thoroughly investigated. Various thresholds are defined on the data being processing by the Hadoop ecosystem. For example, if the energy consumption of a smart home in the smart city exceeds from the normal threshold, an event is generated to the smart community development department for further processing. In addition, both offline and real-time data is tested and processed to get an accurate and fast decision model for future smart city designing. Moreover, the sensor deployment in various buildings, offices, and smart homes is studied and tested to build a generic sensor deployment model. The dataset that is used for the analysis is obtain from various reliable and authentic sources. These datasets include the water usage of various houses in Surrey city, Canada, the intensity of the vehicles at different roads in Aarhus city, Denmark, etc. The details of the data set used in this section are given in Table 1.

TABLE I. DATASET INFORMATION

Sources	Dataset	Size
Surrey city, Canada [8]	Water Consumption	4 MB
Aarhus City, Denmark [9]	Traffic Data	3.04 GB
Aarhus City, Denmark [9]	Pollution Data	77.25 MB

Instead of designing new roads to reduce the traffic intensity, it is better to design a system that can inform drivers about the traffic congestion level on the roads. Therefore, we examine the traffic dataset of Aarhus city, Denmark to build a system that can help drivers to switch to the roads with less traffic. The dataset contains the information of geographical locations of various roads and lanes, average vehicle speed and count at different daytime of Aarhus city. The proposed system efficiently processes the data and generates various events such as traffic intensity warnings when the number of vehicles increases on a particular road from the normal threshold as shown in Figure 2. We set the normal threshold to 8 vehicles on a specific portion of the road. The proposed system generates the warning events in real time and send it to the respective department. The department then broadcast the message to the vehicles headed toward that particular road.

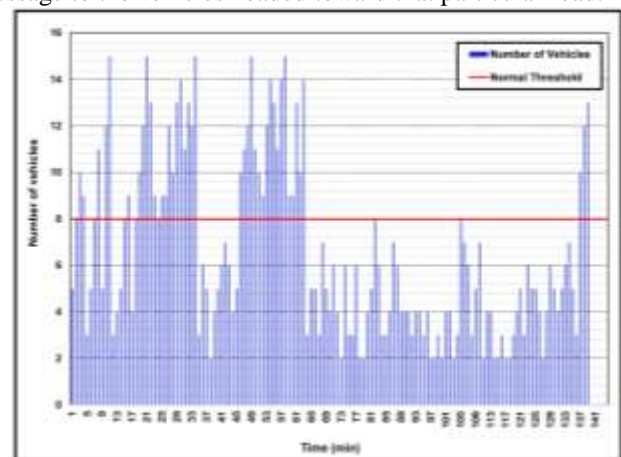


Figure 2 Number of vehicles at different roads in Aarhus, Denmark

The excessive water usage can be a critical problem in near future. Therefore, we analyze the water usage to come up with an appropriate solution for water management. The data set contains the water consumption information of Surrey, Canada. Figure 3 shows that each house consumed more than 80000 to 90000 liters of water each month. This amount of water consumption is very high and therefore in future it can become a serious problem. However, the proposed decision mechanism generates various events to the water management department to take necessary actions to control the water consumption. The warning event is generated when the water consumption is exceeds from 80000 liters. Moreover, new water generations methods can be developed to fulfill the requirements of the citizens.

As the number of factories and vehicles is increasing, the waste production and pollution rise dramatically. Thus, controlling waste management and pollution is becoming a critical issue. In order to design a solution to handle this issue, we analyzed the pollution data of Aarhus city at various time of day. As shown in Figure 4, the quantity of the Ozone ( $O_3$ ) is particularly high at different time of the day. The decision system generates various events to the weather and forecast and health department to circulate a message with the citizen to take great care while visiting the polluted areas.

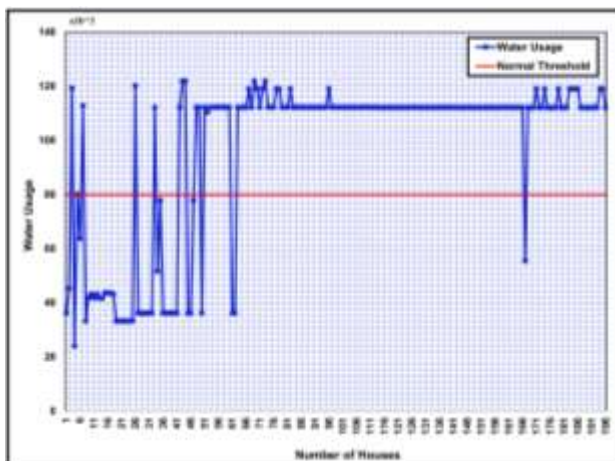


Figure 3 Water usage at different houses of Surrey, Canada

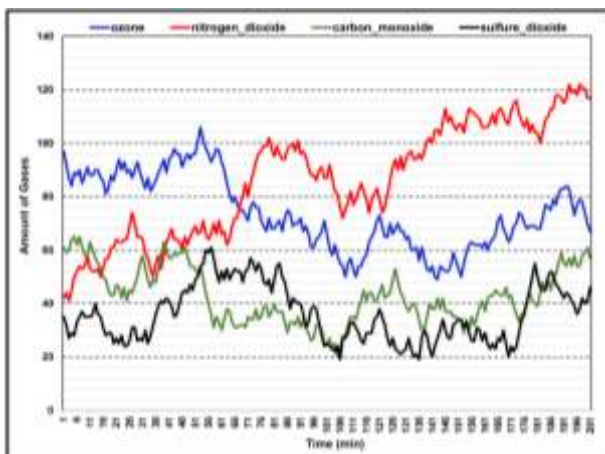


Figure 4 Amount of pollution at different time of day in Aarhus, Denmark

## IV. Conclusion

The extensive expansion of IoT has encouraged the urban networks to be smarter, coining the notion of “Smart Cities”. However, the realization of the smart city is still emerging, since the transformation of the conventional city operations require novelty, networking, processing ability of voluminous data, etc. Therefore, the researchers and industrial experts are keen on shaping a baseline architecture for a realistic smart city.

In this paper, we proposed an architecture of the smart city based on Big Data analytics. The key concern of the study is to ensure intelligent decision management and control center, which mediates the data acquisition sources and applications. The proposed scheme integrates Big Data analytics to facilitate large data processing. Multiple technologies are used on top of the Hadoop storage to facilitate the analysis and decision making. Finally, a real world data set of Surrey (Canada) and Aarhus (Denmark) cities are analyzed to derive the threshold values. In this study, we conceptually proposed the three fold smart city architecture for real-time decision making. In future endeavors, we plan to carry out a simulated experiment to confirm the accuracy and efficiency of the proposed framework.

## Acknowledgment

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2016R1D1A1B03933566).

This study was supported by the BK21 Plus project (SW Human Resource Development Program for Supporting Smart Life) funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (21A20131600005).

## References

- [1] A. Zanella, N. Bui, A. Castellani, L. Vangelista and M. Zorzi, "Internet of Things for Smart Cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22-32, 2014.
- [2] B. Cheng, S. Longo, F. Cirillo, M. Bauer and E. Kovacs, "Building a Big Data Platform for Smart Cities: Experience and Lessons from Santander," in *2015 IEEE International Congress on Big Data (BigData Congress)*, New York, 2015.
- [3] L. Sanchez, L. Muñoz, J. Galache, P. Sotres, J. Santana, V. Gutierrez, R. Ramdhany, A. Gluhak, S. Krco, E. Theodoridis and D. Pfisterer, "SmartSantander: IoT experimentation over a smart city testbed," *Computer Networks*, vol. 61, no. 2014, pp. 217-238, 2013.
- [4] G. Santucci, "Internet of the future and internet of things: what is at stake and how are we getting prepared for them?," in *eMatch conference*, Oslo, 2009.
- [5] T. Ojala, "Open urban testbed for ubiquitous computing," in *Proceedings of Conference on Communications and Mobile Computing*, Shenzhen, 2010.
- [6] M. Zorzi, A. Gluhak, S. Lange and A. Bassi, "From today's INTRANet of things to a future INTERNet of things: a wireless- and mobility-related view," *IEEE Wireless Communications*, vol. 17, no. 6, pp. 44 - 51, 2010.

- [7] T.-h. Kim, R. C. Wai-Chi Fang, S. Mohammed, O. Gervasi and A. Stoica, "Ubiquitous Sensor Networks and Its Application," *International Journal of Distributed Sensor Networks*, vol. 2012, 2012.
- [8] "Dataset Water meters," [Online]. Available: <http://data.surrey.ca/dataset/water-meters>. [Accessed 12 January 2016].
- [9] S. Kolozali, M. Bermudez-Edo, D. Puschmann, F. Ganz and P. Barnaghi, "A Knowledge-Based Approach for Real-Time IoT Data Stream Annotation and Processing," in *IEEE International Conference on Internet of Things (iThings)*, Taipei, 2014.