Smart Urban Planning using Big Data Analytics based Internet of Things

Muhammad Babar, Fahim Arif
National University of Sciences and Technology
Islamabad, Pakistan
babar.phd@students.mcs.edu.pk, fahim@mcs.edu.pk

ABSTRACT
The extensive growth of the Internet of Things (IoT) is providing direction towards the smart urban. The smart urban is favored because it improves the standard of living of the citizens and provides excellence in the community services. The services may include but not limited to health, parking, transport, water, environment, power, and so forth. The diverse and heterogeneous environment of IoT and smart urban is challenged by real-time data processing and decision-making. In this research article, we propose IoT based smart urban architecture using Big Data analytics. The proposed architecture is divided into three different tiers: (1) data acquisition and aggregation, (2) data computation and processing, and (3) decision making and application. The proposed architecture is implemented and validated on Hadoop Ecosystem using reliable and authentic datasets. The research shows that the proposed system presents valuable imminent into the community development systems to get better the existing smart urban architecture.

Author Keywords
Smart City, IoT, Big Data Analytics

INTRODUCTION
The initiative of connecting everyday objects via internet has been progressed with the extensive growth of the smart devices. This growth introduced the idea of Internet of Things (IoT) as the back bone of the web in recent time [1, 4]. IoT has been turned out to be the focal point of the researchers due to the improvement of IoT devices and a rapid growth that is resulting in applications of smart community [5–7]. The idea of smart urban came into being in order to increase the quality of services (QoS) available to the citizens using public resources and services efficiently [8]. In fact, the services are optimized by means of the autonomous data collection including smart home, healthcare, parking, transportation, electricity etc. It is central to process huge quantity of data in real-time to provide the optimized services to the citizens and the traditional analytics and processing may not be suitable for huge data. Therefore, the Big Data analytics integration is first step in the direction of a smarter urban. In addition, there are several individual works presented to cover parking management, water management, garbage management, and so forth [9–10].

Thereupon, computations of the huge amount of data become a need and the urban IoT is incorporated with Big Data analytics. For instance, cameras installed on roads collects the traffic information which is compared with defined values called threshold based on real time processing. Concurrently, the citizen is informed regarding the traffic information to avoid congestion. However, there are a many challenges faced by Big Data and IoT for smart urban. The use of heterogeneous devices brings interoperability issues and Big Data including different formats. Similarly, a number of anomalies are found in the raw data and it is essential to remove the data anomalies the before data processing [11-12]. Several individual are proposed, but still missing a generic and efficient architecture. The realization of smart urban is still in its early years, however, the explicit area such as lighting, smart home, water management, and traffic congestion is focused by different researchers. Therefore, there is a certainty of practical smart urban architecture to have intelligent decisions for the excellence of urban services.

In this research article, Big Data analytics are incorporated with the smart urban to propose a practical architecture for smart cities. The proposed architecture is capable of autonomous data aggregation, real-time intelligent decision-making and user-centric event management. Though, the decision management is the most prominent element for the apprehension of a smart city and the accomplishment of real-time and rapid decisions has become the major goal of the proposed scheme. Hadoop is used for the computation and processing of heterogeneous data in this research. The processing is followed by intelligent decisions generation associated with the smart city operation and events corresponding to the decisions are executed.

The remaining paper is ordered as follow: Section 2 describes recent literature. Section 3 provides the description of the proposed architecture. The results are shown in Section 4. The conclusion is given in Section 5.

UbiComp/ISWC'17 Adjunct, September 11–15, 2017, Maui, HI, USA

ACM ISBN 978-1-4503-5190-4/17/09...$15.00
https://doi.org/10.1145/3123024.3124411

LITERATURE REVIEW

The rapid development in the area of smart urban attracts the researchers and industries in the course of a skilled smart urban design. In current times, researchers are working on a different solution to illustrate the generic design for smart urban using Big Data analytics based IoT. Similarly, a number of simulations based proposals are found in the literature to conquer the issues. The SmartSantander test bed in North Spain is used in [13, 16], where the study allied to the season, temperature, traffic, and working days to facilitate the citizens. Likewise, smart city architecture is also proposed from a data perspective [14]. In addition, 3-tier architecture is proposed to have smooth communication among heterogeneous devices that are connected across ubiquitous platform [15]. A system has been proposed to build up the physical execution of a large-scale IoT infrastructure in a Santander city, [16]. The Hadoop Ecosystem uses MapReduce method for data processing that carries work parallel processing. It works in two different phases: (1) mapping process to transform one dataset values to another set of data, (2) Reduce processes which combines the data and results and reduced in quantity. MapReduce paradigm produces many datasets for the extensive diversity of real world tasks [17]. In the paradigm the data is divided into small lumps and they are dealt in a completely parallel way. In the literature, it is stated about Things that are interconnected and communicate via internet and IoT and social network are associated with each other. To be very cautious they understood this vision as 'Ubiquitous IoT' and it is near to the notion of social IoT [18-19].

Various designs are proposed for demonstrating the integration of social network and IoT in the last decade [20-21]. In the literature, a number of proposals are given to grip a massive data and give IoT services. Although big data has many challenges, however, IoT start to be valuable study [22]. IoT exploits a number of important big data tools for analyzing big data for instance MapReduce, NoSQL, Cassandra, etc. Moreover, different proposals are used to handle offline and online enormous data coming from IoT. Such data can analyze and store using a various storage mechanisms such as cloud and big data techniques which enhance data scalability, availability, flexibility, and adaptability [23]. Processing huge data in the least possible time and decision-making is a very challenging task. Therefore, the existing frameworks are for offline data processing. However, real-time on-line decision-making on huge data is still a difficult task.

The literature depicts a number of challenges that require to be handled, that is, real-time Big Data analytics, aggregation and preprocessing, and a generic model for communication in a smart urban. Therefore, we determine the requirements for a resourceful communication model for smart urban architecture based on Big Data analytics. Big Data analytics involves a number of varied phases including data acquisition and recording, data cleaning, data aggregation, data integration, data representation, data querying, and data analysis. Every phase carries challenges which are needed to be handled.

- Data format challenges found because the big data can be in different shape, semantic, type, and representation.
- Separation of the valuable and helpful data from the noisy data is challenging task and it is very important to filter the data properly.
- The data collection from diverse IoT devices may be loosely controlled. This may bring the resulting values out-of-range, and some impractical data collection may happen. There will be deceptive results, if this issue is not taken into consideration in the architecture of smart urban
- Data heterogeneity is another challenge because isolation carries the intricacy which needs be aggregated properly.
- Missing data refer to the incompleteness may also be found in the data collected from the IoT devices and it generates uncertainties through analysis.
- Timely processing of data is also very important to perform real-time data processing. Since the amount of data increases, it utilizes extra time to evaluate the data and provide immediate results.

PROPOSED ARCHITECTURE

In this part, we talk about the proposed tiered architecture in detail. The purpose of the proposed architecture is to carry out proficient decision making for smart urban. The foundation of proposed scheme is threefold that is data aggregation, data computation, and decision-making. The proposed system is connected to smart urban development societies. Therefore, the data sources include smart traffic, smart weather, smart health, smart home, smart waste, smart water, and so forth. These societies are powered by various wireless communication technologies such as ZigBee, Wi-Fi, Bluetooth, and 3G/4G cellular networks. The proposed architecture is 3-tier (1) data acquisition and aggregation tier, (2) data computation and processing tier, (3) decision-making and application tier. The overview of proposed architecture is depicted in Figure 1.

Data acquisition is performed by smart urban development departments to install sensors in the city. The aggregation is performed using divide-and-conquer method. The pre-processing is carried out to remove and resolve the issue such as out-of-range, impractical data, and missing values. The pre-processing is performed using Min-Max technique of normalization. Similarly, data filtration is performed using Kalman filter to remove noise. The big data processing is carried out via Hadoop Ecosystem using Hadoop two nodes' cluster. Moreover, to realize real time processing, the least slack time (LST) algorithm is used to achieve. In addition, the Round Robin (RR) load balancing technique is also integrated with SLT for time sharing. The processed data is stored and maintained using HDFS storing mechanism. The processed data is separated and sub services are selected properly. The decision making and planning is performed using processed data with the proper decision and event management. The detailed description of each tier of the proposed architecture is given in the next sections.

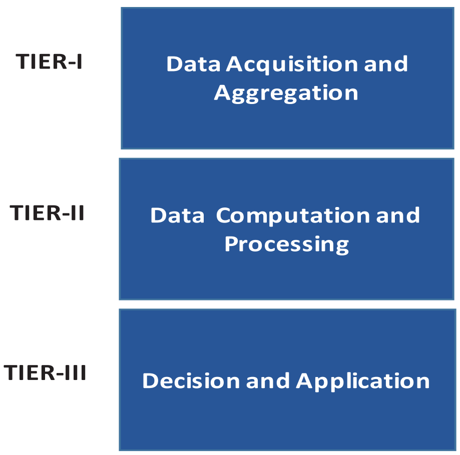

Figure 1. Overview of proposed architecture

Data Acquisition and Aggregation Tier

This tier is responsible for data acquisition and aggregation which are prior to the core processing and depicted by figure 2. Data acquisition is used to sample the signals (real-world situations measurement) and transform to the digital values to be operated by computer (digital machine). The acquisition tasks are performed by various acquisition systems because data acquisition turned out to be a very difficult due the production of the massive amount of data by citizens and connected devices. Therefore, the understanding of the proposed architecture kicks off with the inclusive data acquisition. It is worth mentioning here that, we suppose that the data is acquired by connected societies. The connected societies provide the datasets of their consequent centers (e.g., S_1, S_2,… S_N). In addition, every set is containing K nodes (e.g., NM_1, NM_2 … NM_K), mathematically can be shown as:

$$\text{Data} = \sum_{i=1}^{n} S_i\,, \ \text{where}\ \ S = \sum_{i=1}^{k} NM_i$$

In view of the fact that the process of data acquisition is out of this research, the very initial step of the proposed architecture is to perform data aggregation. Data aggregation is a process which is used to present the grouped data in a summarize form for further analysis. Data aggregation usually applied on huge gigantic big data that do not offer many valued data. The basic idea of data aggregation is to get a hold of more data about groups (i-e healthcare, traffic) based on explicit variables (i-e heart rate, body temperature).

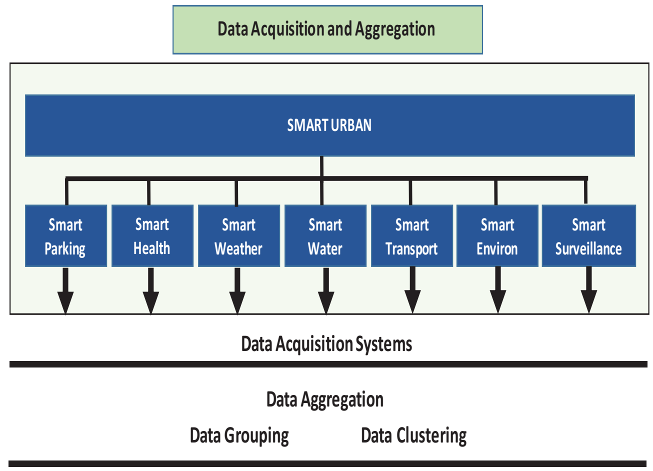

Figure 2. Tier-I of proposed architecture

In this research, the divide-and-conquer technique is used to aggregate the collected data properly. The aggregation process makes use of divide-and-conquer method usefully. The datasets (S_1, S_2, … S_N) first divided into different data blocks (B_1, B_2, … B_n), then the data of the blocks are recognized and classified according to their physical sources (e-g, sensor etc). Then, the data from physical sources are aggregated based on their attributes (explicit variables) mathematically can be as:

$$\text{Set} = \sum_{i=1}^{n} B_i\,, \ \ \text{where every block B} = \sum_{i=1}^{n} Source_i$$

Data Computation and Processing Tier

This tier plays a vital role in this architecture as it is the core processing module. This tier is composed (1) data pre-processor, (2) data filter, (3) Hadoop echo system, (4) and storage. The overall working of this tier is shown in Figure 3. Data collection is usually slackly controlled, unreasonable data union (e-g., patient-sex: male, pregnant: yes), resulting out-of-range values (e-g., water utilization liters), some missing values etc. Therefore, data pre-processing (also known as normalization) is applied before processing. In this research, Min-Max Normalization technique is used to normalize a specific value X to another value Y which fits in the range [m-n]. The task of the filtration process is to filter the data and eliminate noise. We use the Kalman filter (KF) to speed up the data processing and separate the valuable and noisy data. The un-important data is considered noise that does not have an effect on the real-time decision-making and processing of the data. For instance, the information regarding the animals, traffic control police vehicles etc. on the roads from traffic data.

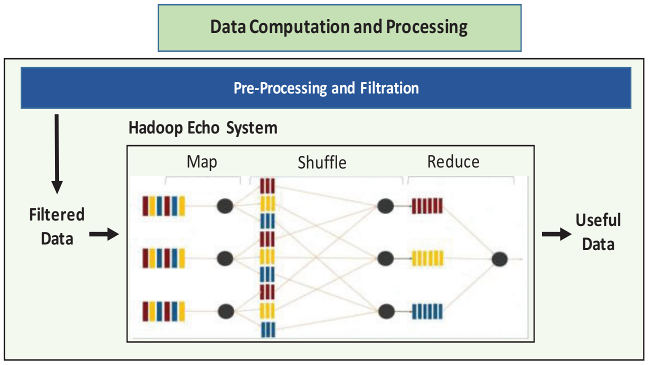

Figure 3. Tier-II of proposed architecture

The Hadoop Ecosystem is responsible for the actual processing and this is the major part of architecture. This server processes and store vast datasets in distributed form and performs parallel processing. The Hadoop echo system gets different datasets (e.g., NM_1, NM_2 … NM_K) and creates results (e.g., DM_1, DM_2 … DM_K) such as:

$$\sum_{i=1}^{n} NM_i => \sum_{i=1}^{n} DM_i$$

The Hadoop Ecosysm uses MapReduce method for data processing that carries work parallel processing. It works in two different phases: (1) mapping process to transform one dataset values to another set of data, (2) Reduce processes which combines the data and results and reduced in quantity, pictorially represented as figure 4.

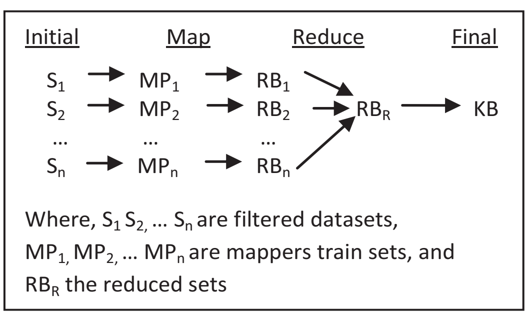

Figure 4. MapReduce Mechanism

The Hadoop performs offline/static processing, therefore, to pull off the real-time processing, we take on a Hadoop two nodes' cluster for and use least slack time (LST) algorithm. First, LST selects those tasks having smallest "slack time (ST)" which can be define as the temporal variation between the time limit, the run time, and the ready time. Considering a process P_i has a deadline D_i, execution time E_i, and released at time R_i, then at time $t > D_i$, the remaining execution time T_{rem} and slack time ST can be defined as:

$$T_{rem} = E_i - (t - R_i), \text{ and}$$

$$ST = D_i - t - T_{rem}$$

The storage device is intended to stock up the results after the processing which are used by decision-making server later. As data storing and processing play a vital role in the comprehension of a smart city, therefore, the proposed architecture makes use HDFS to make possible the data storing and processing easily. The storage requirement of the proposed smart city architecture is assisted by HDFS, which is the main storage of Hadoop. Since the storage of HDFS is distributed, it supplements the MapReduce implementation on smaller subsets of larger data cluster.

Decision-Making and Application Tier

This is the top tier of the propose architecture. The responsibility of this tier is to perform the decision-making and event management. The decision management with and event management is used to generate the decisions and classify the events. The decision-making server describes the decision according to an ontology that is used to unicast the events. The corresponding society performs service selection procedure to distinguish the high and low events. The service selection unit generates the respective event and broadcasts to the implanted notification component which includes departmental, services, and subservices level. Fig 5 represents the overall working of this tier. Let us suppose the traffic data is processed using proposed architecture and the traffic congestion is found on the road at a specific lane A. The decision is taken and event is produced and sent to road congestion control department of the smart traffic society to notify the corresponding users to opt a particular identified lane Y by system to avoid and control congestion.

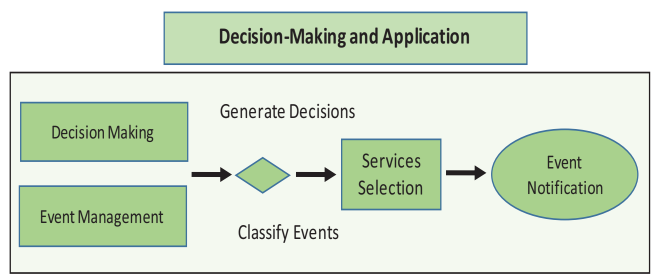

Figure 5. Tier-III of proposed architecture

ANALYSIS AND RESULTS

The implementation is carried out with corei5 processor with 8GB RAM using Hadoop two nodes' cluster on Ubuntu (4GB RAM is devoted to each node) and MapReduce with Java programming. The authenticated and reliable datasets are acquired to validate the proposed architecture. They are: (1) the transportation and vehicular data that is the number of vehicles on different roads in Aarhuscity, Spain [24], and (2) The parking lot dataset includes various parking lots information of Aarhuscity, Denmark generated between May to Nov 2014. Moreover, this data is licensed under Creative Commons Attribution 4.0 International License. Furthermore, the water consumption data of smart homes of the survey, Canada, is gained from the meter readings of around 61263 houses [25]. Water usage data are enclosed by Open Government License of City of Surrey, Canada. In is worth mentioning here that the acquired datasets comprised of the elements of several additional perilous materials but entries of those materials are filtered out for performing real-time decisions.

The center of analysis is to evaluate the proposed work based on already specified thresholds. The threshold for traffic congestion is 9 vehicles per lane. Every time, the data amount at a specific time goes beyond the normal threshold, a particular event is initiated to the relevant department. For instance, the traffic data is processed using proposed architecture and spawns proper events when the number of automobiles exceeds the threshold limit. Figure 6 demonstrates this scenario.

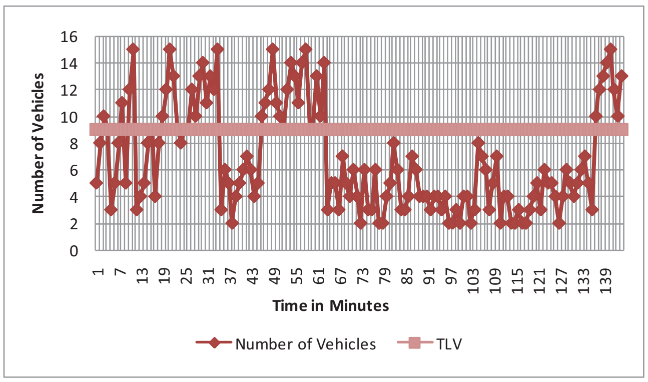

Figure 6. Vehicles on Roads

Similarly, the citizens can make sure the parking space in the surrounding parking lots before driving to the place. Figure 7 shows an option of unfilled spaces in the parking lots during different times. The malls, organizations, and stores parking data can be assessed using the proposed system to offer the smart parking service to the society. In addition, the reservation of the parking lot can also be done before traveling.

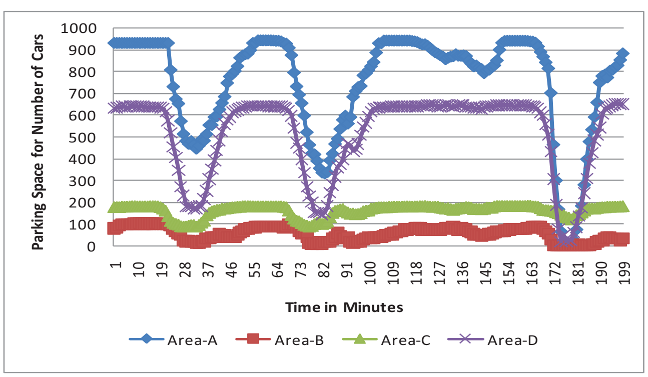

Figure 7. Parking Space in Parking Areas

Similarly, we assessed the usage of water in the city of Surrey, Canada that includes the water consumption. Figure 8 show that every home consumed more than 80000 to 90000 liters of water every month. The threshold for the water consumption is 82000 liters per month. The amount of water consumption exceeding the threshold will cause generating event to the water management department.

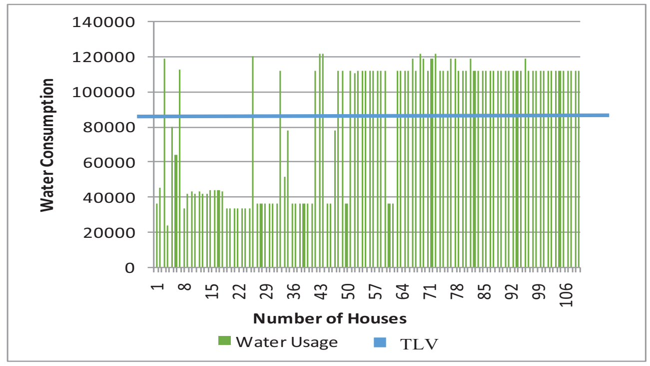

Figure 8. Water Usage in the Surrey City

CONCLUSION

The understanding f the smart urban is still very infant sue to the revolution of the conventional city functions. The smart urban notion brings the researchers and industries to the point to have well-organized and generic architecture. In this research article, an efficient architecture of smart urban is proposed using Big Data analytics. The Big Data analytics is performed using the Hadoop server with MapReduce mechanism, integrated with dynamic scheduling practices. Nevertheless, the proposed scheme does not reflect a generic way out; it is deliberated for explicit objectives. This research intends open concerns of smart urban to make possible the complicated environment for analyzing real time data. The data filtration and normalization are utilized to reduce the processing effort of Hadoop on unsuitable and inappropriate data. Many expertise are utilized on top of Hadoop to make trouble-free analysis and decision-making. At the end, the dataset of Aarhus (Denmark) and Surrey (Canada) cities are examined and evaluated with defined thresholds. In this study, we logically proposed the threefold smart urban architecture for real-time decision-making. In future, we are planning to evaluate the proposed system in the perspective of standardization of smart urban.

REFERENCES

1. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, “Internet of things: a survey on enabling technolo-gies, protocols, and applications,”IEEE Communications Surveys and Tutorials,vol.17,no.4,pp.2347–2376,2015.
2. Awais Ahmad, Anand Paul, M. MazharRathore, and Hangbae Chang. "Smart cyber society: Integration of capillary devices with high usability based on Cyber–

Physical System." *Future Generation Computer Systems*56 (2016): 493-503.

3. Guinard, Dominique, et al. Towards physical mashups in the web of things. Networked Sensing Systems (INSS), 2009 Sixth International Conference on.IEEE, 2009.
4. M. A. Razzaque, M. Milojevic-Jevric, A. Palade, and S. Clarke, "Middleware for internet of things: a survey,"*IEEE Internet of Things Journal*,vol.3,no.1, pp.70–95,2016.
5. J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An informa-tion framework for creating a smart city through internet of things,"IEEE Internet of Things Journal,vol.1,no.2,pp.112–121, 2014.
6. S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: a comprehensive survey,"IEEE Access, vol. 3, pp. 678–708, 2015.
7. M. Khan, B. N. Silva, and K. Han, "Internet of things based energy aware smart home control system,"IEEE Access,vol.4, pp. 7556–7566, 2016.
8. A.Zanella,N.Bui,A.Castellani,L.Vangelista,andM.Zorzi , "Internet of things for smart cities," IEEE Internet of Things Journal,vol.1,no.1,pp.22–32,2014.
9. R. E. Barone, T. Giuffr` e, S. M. Siniscalchi, M. A. Morgano, and G. Tesoriere, "Architecture for parking management in smart cities,"IET Intelligent Transport Systems,vol.8,no.5,pp.445–452, 2014.
10. M. Khan, S. Din, S. Jabbar, M. Gohar, H. Ghayvat, and S. C. Mukhopadhyay, "Context-aware low power intelligent Smart-Home based on the Internet of things,"Computers & Electrical Engineering,vol.52,pp.208–222,2016
11. B. N Silva, M. Khan, and K. Han, "Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making", "Wireless Communications and Mobile Computing, Vol 2017, Article ID 9429676,
12. Awais Ahmad, Anand Paul, and M. Mazhar Rathore, "An efficient divide-and-conquer approach for big data analytics in machine-to-machine communication", Neurocomputing 174 (2016) 439–453
13. B.Cheng,S.Longo,F.Cirillo,M.Bauer,andE.Kovacs,"Bu ild-ing a big data platform for smart cities: experience and lessons from santander," in Proceedings of the 4th IEEE International Congress on Big Data (BigData Congress '15), pp. 592–599, New Yo r k, NY, US A, J u ly 2 01 5 .
14. W.Rong,Z.Xiong,D.Cooper,C.Li,andH.Sheng,"Smartci ty architecture: a technology guide for implementation and design challenges,"China Communications,vol.11,no.3,pp.56–69, 2014.
15. S. V. Nandury and B. A. Begum, "Smart WSN-based ubiquitous architecture for smart cities," inProceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '15), pp. 2366–2373, Kochi, India, August 2015.
16. L.Sanchez,L.Mu˜ noz, J. A. Galache et al., "SmartSantander: IoT experimentation over a smart city testbed,"Computer Networks, vol. 61, pp. 217–238, 2014.
17. Dean, Jeffrey, and Sanjay Ghemawat. "MapReduce: simplified data processing on large clusters." Communications of the ACM 51, no. 1 (2008): 107-113.
18. S. Uppoor, O. Trullols-Cruces, M. Fiore, J.M. Barcelo-Ordinas, Generation and Analysis of a Large-scale Urban Vehicular Mobility Dataset, IEEE Transactions on Mobile Computing, Vol.13, No.5, May 2014
19. Ning, Huansheng, and Ziou Wang. "Future Internet of things architecture: like mankind neural system or social organization framework?." Communications Letters, IEEE 15, no. 4 (2011): 461-463.
20. M. Mazhar. Rathore, A. Paul, A. Ahmad, B.W. Chen, B. Huang, Wen Ji "Real-Time Big Data Analytical Architecture for Remote Sensing Application", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume: PP, Issue: 99, Page(s): 1 – 12, Digital Object Identifier 10.1109/JSTARS.2015.2424683, 1939-1404 © 2015 IEEE.
21. Chung, T.Y., Mashal, I., Alsaryrah, O., Chang, C.H., Hsu, T.H., Li, P.S. and Kuo, W.H., 2014, September. MUL-SWoT: A Social Web of Things Platform for Internet of Things Application Development. In Internet of Things (iThings), 2014 IEEE International Conference on, and Green Computing and Communications (GreenCom), IEEE and Cyber, Physical and Social Computing (CPSCom), IEEE (pp. 296-299). IEEE.
22. Labrinidis A, Jagadish HV "Challenges and opportunities with big data", Proc VLDB Endowment 5(12):2032–2033, 2012.
23. Le, Xuan Hung, Sungyoung Lee, Phan Tran Ho True, Asad M. Khattak, M. Han, D. V. Hung, Mohammad M. Hassan et al. "Secured WSN-integrated cloud computing for u-life care." In Proceedings of the 7th IEEE conference on Consumer communications and networking conference, pp. 702-703. IEEE Press, 2010
24. T. Dataset, Dataset Collection, http://iot.ee.surrey.ac.uk:8080/datasets.html#traffic.
25. Dataset Water meters, http://data.surrey.ca/dataset/water-meters