IoT-Based Big Data: From Smart City towards Next Generation Super City Planning

Chapter · January 2020

DOI: 10.4018/978-1-5225-9866-4.ch064

4 authors, including:

Muhammad Mazhar Ullah Rathore
University of New Brunswick

Anand Paul
Kyungpook National University

Awais Ahmad
Air University

International Journal on Semantic Web and Information Systems
Volume 13 • Issue 1 • January-March 2017

IoT-Based Big Data:
From Smart City towards Next Generation Super City Planning

M. Mazhar Rathore, School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea
Anand Paul, School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea
Awais Ahmad, School of Computer Science and Engineering, Kyungpook National University, Daegu, South Korea
Gwanggil Jeon, Department of Embedded Systems Engineering, Incheon National University, Incheon, South Korea

ABSTRACT

Recently, a rapid growth in the population in urban regions demands the provision of services and infrastructure. These needs can be come up wit the use of Internet of Things (IoT) devices, such as sensors, actuators, smartphones and smart systems. This leans to building Smart City towards the next generation Super City planning. However, as thousands of IoT devices are interconnecting and communicating with each other over the Internet to establish smart systems, a huge amount of data, termed as Big Data, is being generated. It is a challenging task to integrate IoT services and to process Big Data in an efficient way when aimed at decision making for future Super City. Therefore, to meet such requirements, this paper presents an IoT-based system for next generation Super City planning using Big Data Analytics. Authors have proposed a complete system that includes various types of IoT-based smart systems like smart home, vehicular networking, weather and water system, smart parking, and surveillance objects, etc., for dada generation. An architecture is proposed that includes four tiers/layers i.e., 1) Bottom Tier-1, 2) Intermediate Tier-1, 3) Intermediate Tier 2, and 4) Top Tier that handle data generation and collections, communication, data administration and processing, and data interpretation, respectively. The system implementation model is presented from the generation and collection of data to the decision making. The proposed system is implemented using Hadoop ecosystem with MapReduce programming. The throughput and processing time results show that the proposed Super City planning system is more efficient and scalable.

KEYWORDS

Big Data, Hadoop, IoT, Smart City, Super City

1. INTRODUCTION

According to the report published by CISCO (2008), things connected to Internet surpassed the number of people living on earth. It is also added that these things will touch the limit of fifty billion by 2020, taking us to the world of digitization. These things interact and communicate with each other with the help of internet – we call it the Internet of Things. Recent advancements in the domain of IoT have aided human lifestyles significantly, traversing fields of healthcare, automation and transportation, and disaster management and response.

DOI: 10.4018/IJSWIS.2017010103

In the domain of IoT where different things are involved in it strengthening the capabilities of different devices. These strengthening capabilities include hearing, seeing, listening and communicating altogether with each other. Thus, IoT transforms those objects from being traditional toward smart by incorporating its ubiquitous and pervasive computing, embedded devices (e.g., actuators, smartphones, tablets, and other networked enabled devices), communication technologies, sensor networks, Internet Protocols and applications, revolutionize the way of human beings. The Internet will be no longer considered as the network of computers. Rather, it will be involved with billions of smart devices along with the embedded systems. As a result, Internet of Things (IoT) will greatly increase its size and scope, providing a new way of opportunities as well as challenges (Zeng, Guo, & Cheng, 2011). The majority of the countries have put forward longstanding national strategies for the implementations of IoT after completing the intangible stage of service level. For instance, Japan's broadband access is providing the facility of communication between people, people and things, and things and things (Srivastava, 2004). Similarly, S. Korea's smart home enables its people to access things remotely (Giroux & Pigot, 2005). Singapore's next generation I-Hub (Han, 2005) intends to comprehend the next generation "U" type network through a secure and ubiquitous network (O'droma & Ganchev 2010). The stated initiatives laid the foundation of IoT (Xia et al., 2012). Also Now IoT tends to the Web of things that's aiming to develop smart societies (Ahmad et al., 2016). Thus, a growing need for standardization for sustainable cities rises (ITU-T Focus Group on Smart Sustainable Cities, 2015). In the literature, extensive research work performed on the smart technology has been observed (Dixit & Prasad, 2007). Similarly, the idea of the smart home is also extended towards the Smart Community and Smart Society where Home Domain, Community Domain, and Service Domain are integrated to provide benefits to mankind.

Due to the incorporation of ubiquitous and pervasive computing, the trend of living has now changed. It is predicted that by 2050, seventy percent of the world population will live in cities (Jin et al., 2014). Moreover, a rapid increase in population migration towards cities has been seen. Therefore, it results in enhancing the number of things to be interconnected with each other over the internet, which results in generating a massive volume of data with heterogeneous properties (termed as Big Data). To enrich the smart technology, such as Smart and Super City, the better analytics of Big Data could play a vital role in the advancement of Information and Communications Technologies (ICTs). Such of Big Data analysis provides a better understanding and useful information about the future as well as planning and development, while showing us a deeper insight of Big Data. Hence, by analyzing such data based on the user needs and demands, the cities can be built smarter. Thus, the enabling technologies and the data analytics made the IoT to come out of its primary phases.

Traditionally, for urbanization, it is an extreme important factor to realize the desires for service profiling to enrich the efficiency and bring the recent advancement in the city administration. Currently, few organization are working on their platforms for live monitoring of various process parameters. Such activities are followed by huge amount of data collection, Data filtration and processing, data analytics, and decision making. Usually, data gathering technique is expensive and hard to achieve. Therefore, it is necessary to integrate smart technologies that could efficiently generate and aggregate the data, perform analysis, and foresee the future for superior planning and development.

With the advancement in IoT technology, smart systems, and Big Data analytics, the world is evolving to be smarter. The modern era is now turning from smart systems to Smart City and then from Smart City to Super City. Super City is the city in which everything is in perfect and in a near ideal condition. We can say, the Super City is the advanced form of Smart City with extra facilities not covered by the Smart City. In addition to the facilities of Smart City where people are facilitated anytime, anywhere for anything, in Super City people are more secure and safe from theft, robbery, assaults and other crimes as well as from external environments such as pollution. Super City has a neat and clean environment, good sanitation system, lot of parking areas at proper places, parking and cycling tracks, and many more. Resultantly, people living in the Super City are intelligent and have good health. Industries, roads, houses; all are constructed in an ideal way at proper locations with suitable structures. Number and location of roads and parking space fulfill the present and future demands of the citizens. The city has no scarcity of resources. All resources needed, such as energy, water etc., are premeasured and managed beforehand. The Super City also has safety from natural disasters, such as floods, tsunami, thunder storms, heavy rains, earthquakes, etc. Even though, we cannot escape from such disasters, however, if any disaster happen, the city has the technology to detect them in advance and take precautionary measures to save the citizens' life. Here, in this paper, authors' main objective is to give directions to the municipalities to use IoT technologies with various perspective to plan and achieve Super City using Big Data analysis techniques in a time efficient manner. IoT generated Big Data can be used to build Smart City by real-time data analysis. Authors' aim is to use the same resources such as smart systems and data generation with the historically stored information for Super City planning. In other words, authors proposed the way, they can use IoT generated Big Data and perform time efficient analysis for Smart City using real-time data and for Super City planning using larger size offline historical datasets.

Having understood the practicability and potential of IoT-based smart systems, in this article, authors drive the concept of Smart City toward the Super City while working with Big Data analytics. The contributions of the work is manifold, and are as follows:

- The complete architecture, with four layers/tiers, is proposed to develop Super City by precise planning using Big Data analytics. The proposed architecture is proficient enough to analyse huge amount of datasets generated from various sources of the IoT-based smart systems in the city i.e., vehicles, road sensors, smart car parks, smart homes, weather and pollution monitoring systems, etc.
- Also, the system implementation model is given with the guidelines to implement the system effectively.
- In addition, using the proposed system, the analysis is performed on the big datasets in order to plan for Super City. These datasets are generated from smart systems, which are deployed in various cities.
- Finally, the system is implemented as a test case using Spark over Hadoop. It is evaluated in terms of efficiency while considering processing time and system throughput on various datasets.

The paper is organized into seven major sections. Section 2 describes the motivation behind their work. Section 3 presents the details of the proposed model and architecture. Section 4 describes the analysis on urban data to make Super City planning. This section includes the details of datasets, analysis and discussion on each type of urban datasets. System implementation details and evaluation is done in Section 5 and Section 6, respectively. Finally, Section 7 concludes their work and gives future directions.

2. MOTIVATION

Even though many Smart City system have been proposed in the literature, but all of them are specific to particular application e.g., traffic control, smart lights in smart city, smart card for transport tickets, automatic number plate recognition (ANPR) systems for intelligent transportation systems (ITS), automatic meter reading (AMR), automated monitoring of public service provision using RFID chips (Dodge & Kitchin, 2007; Hancke & Hancke 2012), mission control system in case of accidents (Singer, 2012), smart electricity networks (Kyriazis et al, 2013; Bulkeley, McGuirk, & Dowling, 2016), smart transportation (Rathore et al, 2016), etc. Therefore, the authors' intention is to enhance the massive distribution of ICT resources in developing the entire system. The advancement in the present technology in the IoT reveals the tendencies of ICT. Thus, a system is desired that could circumscribe all of the most recent progresses in the field of IoT; as a result, a notable development can be realized in the near future. Furthermore, it can be seen that most of the city data analysis are already being performed for city planning. However, in a bulky environment where massive amount of data is being collected, it is obvious to left huge portion of data as disjoint at the time of analysis. Consequently, such analysis does not withdraw healthier understanding of the situation to make future plans. For this reason, Super City planning provide a new way to the field of the IoT in which the data is integrated and analysed based on the device geographic location to attain various services in a city.

Additionally, due to the fact that Smart City applications could take assistance from IoT-based smart systems capabilities that can be categorized into different impact areas (Jin et al, 2014). That can affect the life of a citizen while providing good health and safety, well established smart and intelligent transportation system, safety from pollution, and so on so forth. Various smart projects are being driven using sensors, which are related to monitoring cyclists, cars, public car parking, etc. Apparently, different other service domains have been identified that apply Smart City IoT framework to deliver facilitates like noise and air pollution control, smart vehicle mobility, and city safety system using real-time surveillance in the cities. Still, the current research have a limited research findings in the area of Smart City and urban planning. Similarly, a compact, scalable, and efficient system is still the need of the municipalities. In the recent years, the Big Data is widely being used by the authorities to analyze the different facets of the City. A similar concept can be followed using the IoT paradigm for city data collection and then the use of Big Data analytics for Super City planning. Thus, the authors have tried to come up with a solution that is applicable for the urban area planning. They are intended to apply immediate analytics in the case of real-time Smart City decisions and offline historical datasets analysis for future Super City planning and future enhancements. Their objective is to provide a complete solution that can be used by authorities in order to start their growth from simple city to Smart City and from Smart City towards Super City. Their aim is to have a complete architecture and its implementation model with efficient and fast tools that can effectively and efficiently process the huge amount of historical city data and apply Big Data analytics in order to establish well-planned city.

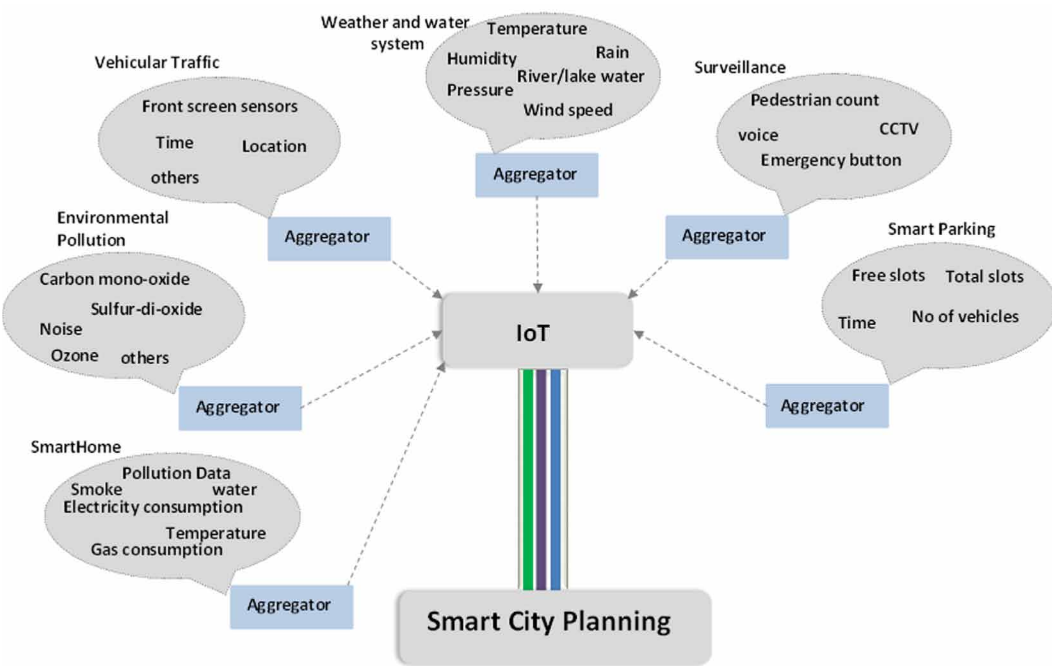
3. PROPOSED SCHEME

In order to generate data for analysis aiming at establishing IoT-based Super City planning system, authors have been proposed to deploy numerous sensors, surveillance cameras, emergency street buttons for safety, and other fixed devices. The data from existing smart systems in the city is also used for Super City planning. But, all sensors and smart systems generate abundant amount of Big Data with very high speed. The Hadoop ecosystem is used to process that data in an efficient way. This section demonstrates the proposed architecture containing the description of how the sensors are deployed and how the sensors generates data. The implementation model of the system is also depicted in the current section.

3.1. Use of IoT for Super City Planning

For next generation Super City planning, the IoT-based smart systems are integrated for city data generation, as shown in Figure 1. That smart systems' data is stored at large storage that can later be analyzed for future Super City planning. By smart home generated data, government authorities can analyze the electricity consumption of the previous year, which gives the prediction of the electricity demand for the coming year. Consequently, municipalities can take the corresponding action to accomplish the future needs and make future plan for building new dams to produce more energy. Moreover, the patterns of the power usage at all houses can also be analyzed with different time periods in order to manage the power bills accordingly to properly assist the citizens. Moreover, various energy plans can also be set for various time-periods of the year subsequently. Also, from smart

Figure 1. Sensors deployment and smart systems data generation



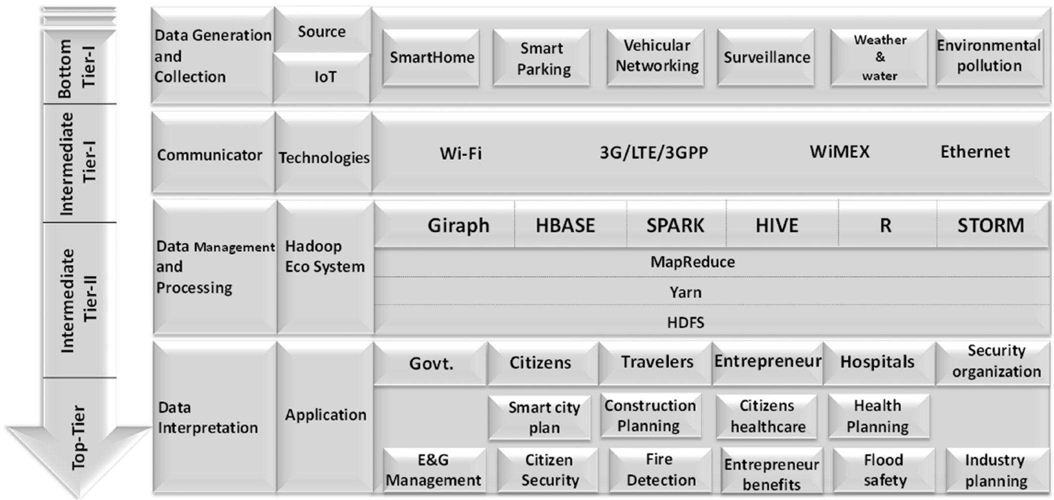
parking and vehicular traffic dataset analysis, the future plan can be designed for new parking lots, new buildings and malls, new roads or extending exiting roads based on future needs. Furthermore, to identify the causes of air-pollution, we need to analyze the environmental data, which can also be used to find out the places with dangerous level of toxic gases. Authorities might made plan to reduce the pollution causes e.g., transferring factories, diverting traffic, etc. similarly, by analyzing the weather and water resources datasets, we can make agriculture related and flood safety plans. Additionally, based on the temperature datasets and city electricity expenditure datasets, we can make a better strategies to save the electricity during the hot seasons. Moreover, historical surveillance data is used to investigate the unsafe places, number of crimes events, type of crime which is on the rise, number of victims, etc. Based on these upshots, the security needy locations for the next year or even for next month can be planned.

3.2. Big Data Analytical Architecture for Super City Planning

Keeping in view the needs of the Super City, authors proposed an architecture to analyze IoT Big Data for future city planning. The complete system architecture is depicted in Figure 2. The analysis system is composed of four Tiers including Bottom Tier, Intermediate Tier-I, Intermediate Tier-II, and the last Top Tier.

Tier 1. Bottom Tier: At this layer, the data is generated using numerous IoT systems and sensors. Later the data is collected and aggregated using a separate aggregator for each type of system. Since there are a lot of various kinds of sensors that are participating in the environment sensing, therefore, which results in the production of enormous amount of heterogeneous data with varying formats, varying origin and varying periodicity. Besides, some data is critical and requires safety, privacy, security, and quality constraints. Also, the raw data collected from sensor always have greater amount of the metadata than the real measure. Hence, at early stages of this layer, the

Figure 2. Proposed architecture for Super City planning



registration and filtration techniques are applied to filter the needless metadata. The repeated data is also discarded.

Tier-II. Intermediate Tier-I: This tier is responsible for providing communication facility between sensors, sensors to relay node by means of ZigBee technology. It also handles the data delivery from relay to gateway node and then from base station towards internet cloud using different communication technologies. At analysis unit with various analysis servers, Ethernet is used.

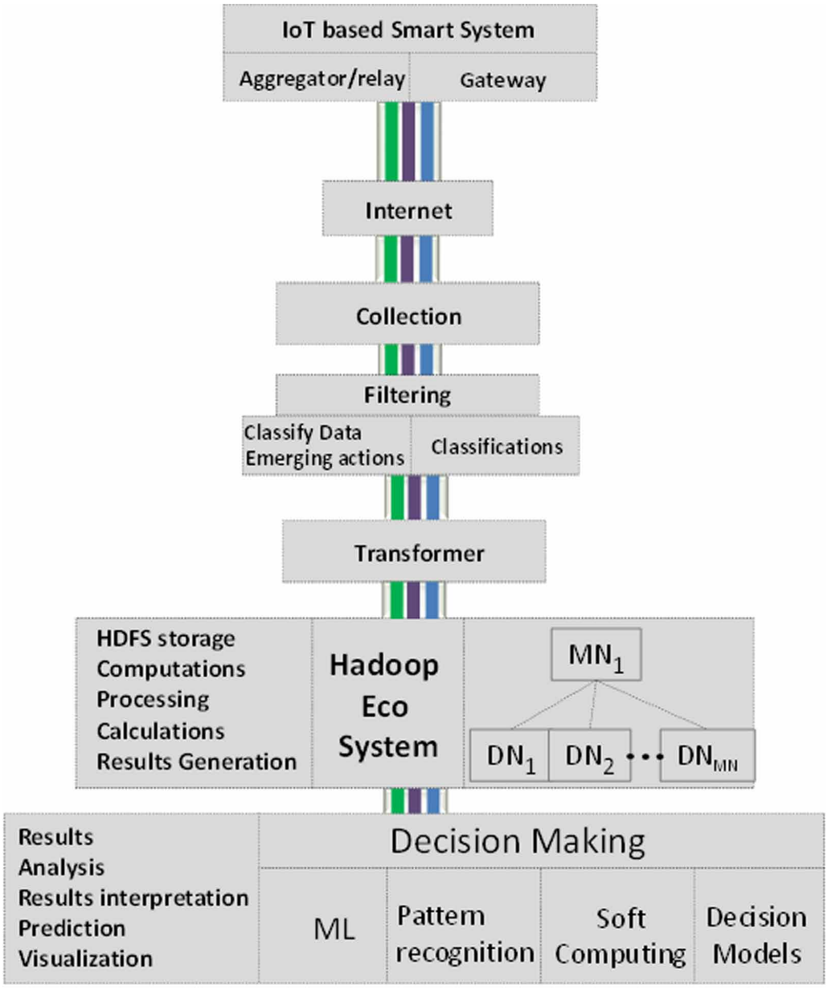
Tier-III. Intermediate Tier-II: This is the primary layer of the entire system handling the data processing part of the system. They proposed to use Hadoop ecosystem for storing (Hadoop Distributed File System (HDFS)) and processing (MapReduce) huge amount of historical data. However, it is also possible to use any third-party tool that can be combined with Hadoop for efficiency or simplicity. However, for system evaluation, they implemented the system by using MapReduce for data processing mechanism and HDFS for data storage. However, it is also sometime convenient to use HIVE, HBASE, and SQL for handling in-memory as well as offline Database to store historical information.

Tier-IV. Top-Tier: The last tier is the Top-Tier that is responsible for the interpretation and the use of results for various applications and decision making. It also generated reports based on results to facilitate the city planners.

The researchers also design a model for system implementation that can guide the authorities to perform urban planning by implementing the proposed system. The model is presented in Figure 3. In the beginning, every smart system/sensors would generate their data by sensing the environment. At every smart system have a relay node connected with it using ZigBee communication. That node aggregates the data from all related systems/sensors. The relay node then transmits the data to the main analytical system by passing through the GW and the Internet. Since majority of the information in the raw data from sensors is metadata, therefore, all the unwanted metadata and repetitive data is discarded. Moreover, the sensor data has the information of message type and identifier that is used to classify the data. Subsequently, the classified data is transformed to the Hadoop readable file format such as, sequence file that is understandable to the Hadoop ecosystem.

Since they are handling with a Big volume of data (named as Big Data), the authors require an efficient system that could speedily process large datasets. To fulfill this basic requirement, the authors deployed Hadoop ecosystem that is equipped with few master nodes (one or more) and numerous

Figure 3. Implementation model for IoT-based big data analysis



data nodes working under a master node. The Hadoop ecosystem stores large datasets using Hadoop Distributed File System (HDFS) by dividing them into an equal size of chunks. Later, it stores these chunks in various data nodes by replicating every chunk on more than one data nodes in order to achieve more safety and reliability. Next, these chunks are processed in parallel in order to perform analysis using MapReduce system. All the chunks' results generated/calculated by each Map function are aggregated using Reduce function on Hadoop ecosystem. Finally, the decision making is performed based on the results generated by Hadoop ecosystem. The decision-making approach uses machine learning, pattern recognition, soft computing and decision models.

4. DATA ANALYSIS FOR SUPER CITY PLANNING

The authors performed analysis on various IoT datasets to execute the feasibility study and the importance of the proposed system for Super City planning benefiting the future generation of citizens. The analysis makes it evident that how the IoT-based smart system can play a major role in Super City planning. It shows how they can use the huge amount of historic sensors data in Big Data analytics for Super City planning. This section describes the datasets used for analysis and evaluation purpose, and the analysis made which clarifies how they use IoT-based smart system for Super City planning.

4.1. Dataset Description

As it is very difficult at this time to implement all the smart systems for data generation, therefore they take already generated datasets through mentioned smart systems from various reliable resources. The datasets includes 1) flood dataset which describes all the flood occurrence in the world, 2) the smart home sensors collected dataset including temperature, water usage of each house, etc., 3) the vehicular network datasets which includes the vehicle's information such as speed, location between various source and destination points as well as the general mobility information of all the vehicles in the city, 4) the parking lot datasets which contains the current usage of the parking area, 5) pollution datasets describing the various toxic gas levels in the air, 6) social media such as Twitter datasets including tweets record, 7) weather datasets including incessant measurement of heat, moisture, humidity, rain, etc., outside as well as inside the home. A brief detail of all the datasets including the area, number of parameters, short description, and the source from where the datasets are taken, is given in Table 1.

Moreover, the social network has major importance in Super City planning. So they take twitter data whichincludes tweets from many people of the city, their location, time, etc. It contains the 13674 tweets records from September 23, 2013, to December 17, 2013. The Twitter data analysis is a broad topic to cover, so they did not perform deep analysis on Twitter. However, they use Twitter data for evaluation purpose.

4.2. Analysis and Discussion

The main challenge in Super City planning is the analysis of the huge amount of IoT data which can be accomplished by parallel processing through Hadoop. Proper Super City planning not only benefits the government authorities but also the public by getting high security, healthy environment, and in a nutshell, a smart and developed city. Here, authors are analyzing various kinds of data and give guidelines to the consultants how they can use IoT technologies and the Big Data produced from IoT for urban development. Using the proposed system and data analysis, the authorities can achieve the Super City for their next generation citizens. They mainly focus on the analysis of vehicular traffic, smart home, parking lots, flood, and pollution.

Table 1. Datasets details

S#	Datasets	Area	No. of parameters	Ref.
1	Water Usage	Surrey Canada	11	(GIS Section, 2014)
2	Madrid Highway vehicular traffic	Madrid	5	(Gramaglia, 2014)
3	Vehicular Mobility Traces	Cologne Germany	5	(Uppoor & Fiore, 2011; Naboulsi & Fiore, 2013; Uppoor et al., 2014)
4	Parking lots	Aarhus Denmark	7	(Bischof et al, 2014; Tönjes et al, 2014; Kolozali et al., 2014)
5	pollution	Aarhus Denmark	8	(Bischof et al, 2014; Tönjes et al, 2014; Kolozali et al., 2014)
6	City traffic	Aarhus, Denmark	9	(Bischof et al, 2014; Tönjes et al, 2014; Kolozali et al., 2014)
7	Weather	Aarhus, Denmark	7	(Bischof et al, 2014; Tönjes et al, 2014; Kolozali et al., 2014)
8	Floods	World	30	(G.R.Brakenridge, 2010)

4.2.1. Vehicular Traffic Analysis

For using vehicular traffic information for urban development, authors take Madrid Highway vehicles data and the data covering the city of Aarhus city of Denmark. This dataset is more significant in Super City planning to facilitate the authorities to build new roads, buildings, etc. The Madrid Highway vehicular traffic data covers the locality and speed of each vehicle on the Madrid highway. They also tested the vehicular mobility dataset which is produced by Institute of transportation systems, German Aerospace Center (ITS-DLR) for TAPAS Cologne project and covers mobility of 700 cars in 400 square km area of Cologne city of Germany for 24 hours. The Aarhus city dataset is produced by placing several source and destination pair sensors on different locations to estimate the traffic between two points. They cover the various aspects of cars such as time and average speed of vehicles.

For Madrid Highway traffic analysis, they show the traffic intensity on first 2550 location of the Highway for a particular time in Figure 4. All can find the location where the intensity of the traffic is more. At the start of the Highway, the traffic intensity is high and then gradually decreases. This shows that the start location, i.e., 500 is the key location where most of the cars are passing through. On the other hand, the location 2500 is very far from the city, where very few cars are traveling. This shows we need to build more lanes at the start of the highway and few lanes at the end to properly handle the intensity of traffic with optimum resources. We can also plan to build new roads when the intensity of traffic is too much on the highway. Moreover we can also assume that at location 2500, the number of people living or the number of houses, shops, and building are less, therefore we can plan to build more houses and buildings there to reduce the traffic burden, pollution, and crowd too.

For the same Highway, the speed analysis shows the average speed of the vehicles is 90 km/h which is quite high. However, by speed analysis we can evaluate the road condition; where the average speed is lower at maximum time, the condition of the road is not good, or there is more crowd at that place. Therefore, we can decide to repair or build new roads based on the speed analysis as well. Moreover, authors also performed over-speed analysis on Madrid highway. The location where most of the vehicles crosses over-speed limit is shown in Figure 5. This analysis helps the authorities to place sign boards and the sergeants to control the speeding of cars thereby avoiding accidents. We

Figure 4. Location and the intensity of vehicles on Madrid Highway

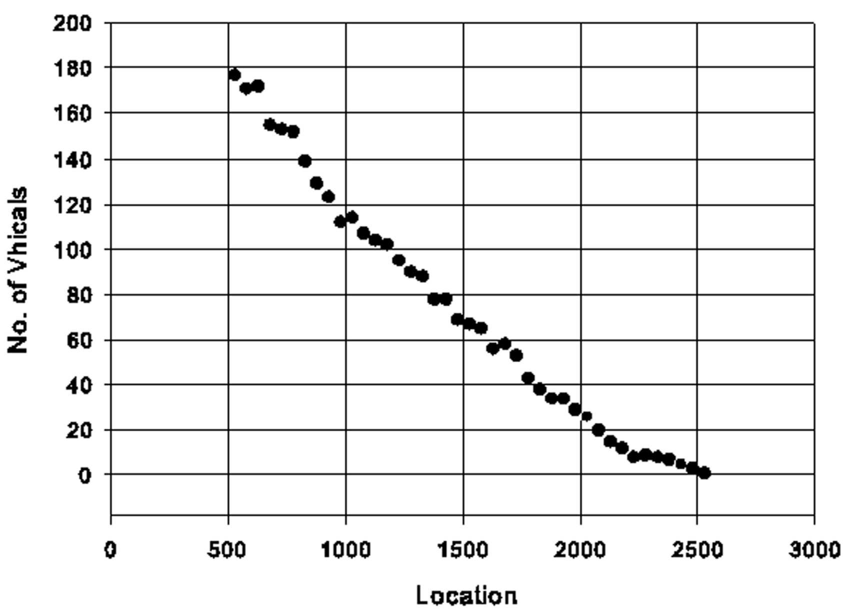
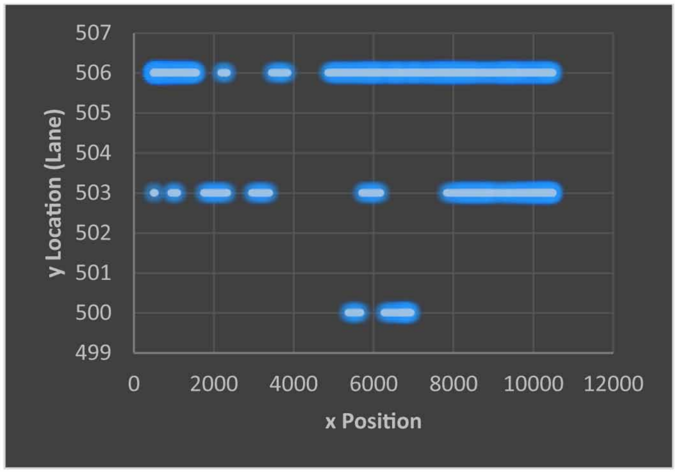


Figure 5. Location of over-speed vehicles on Madrid Highway



can see more vehicles crosses the maximum limit from location 5000 to 11000, so it is a dangerous location. Most of the violations occur at lane three (most extreme lane) of the highway. These violations might be due to the less number of vehicles on that lane. This can be stopped by notifying through sign boards or placing speed barkers at that location, which is suitable for that location. Moreover, it is also a better option to monitor the accident ratio with respect to the speed on the highway.

The intensity of traffic analysis with respect to time in Aarhus city is performed as seen in Figure 6. We can perceive from time 08:25 and 11:55, the intensity of some vehicles is higher i.e. >12. Thus, a well-organized road system should be planned that can dynamically adjust the routes during the peak hours. Also, Figure 7 shows the analysis result of the speed with respect to some vehicles. We can understand that if the vehicle speed is low on some roads then the intensity of the cars is higher on that road such as at time 106 and 121, the vehicle speed is quite low i.e., 45 and 42. Thus, such statistics can be used to develop wider roads where the intensity of vehicle is higher and vice versa while planning for future.

4.2.2. Smart Home Data Analysis

Smart Home data analysis also played a vital role in Super City planning including the energy management, water management, house building structure planning etc. As it is a very broad topic

Figure 6. No. of vehicles on the road at various time

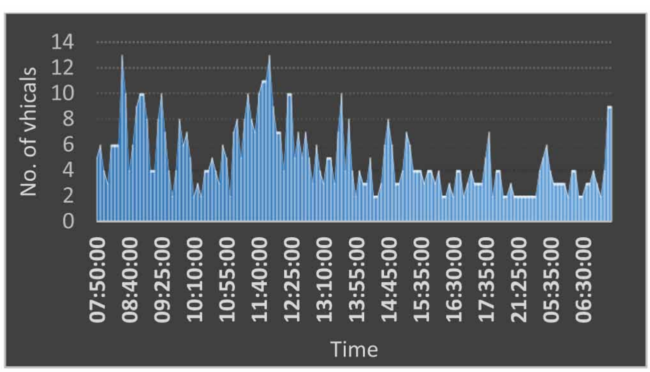
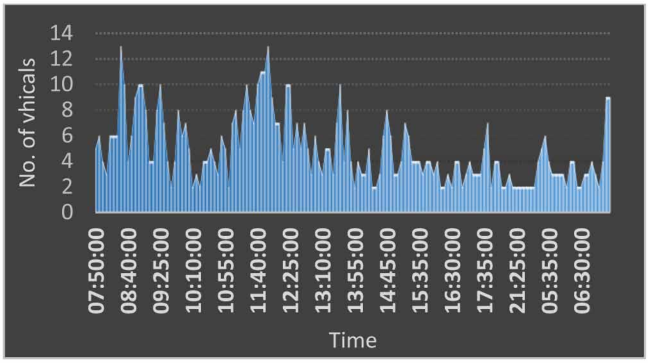


Figure 7. Relationship between number of vehicles and the average speed



to cover, here, the authors are only presenting the analysis with respect to the water management and planning based on individual house usage of Surrey city of Canada. The dataset consists of 61263 houses water meter readings and complete address and reading period of the house.

From the water analysis as shown in Figure 8, which displays the normal use of water at each house, we can predict the next month's or next year's need of water. The analysis shows that more than 6000 houses consume more than 8000-9000 cubic meter of water, which is the estimation of the normal use of water at a maximum number of houses. Figure 9 shows the area wise usage of water from which authorities can control the flow of water depending on the need of the area. For instance, 101B and 102B have low water consumption and 102 and 103A have very high water consumption. To check the validity of the statistics as per water usage and predictions, authors use the skewness statistical measurement with the household water consumption W_1, W_2, W_3 . . . W_N by the following equation.

Figure 8. Water usage count of the city

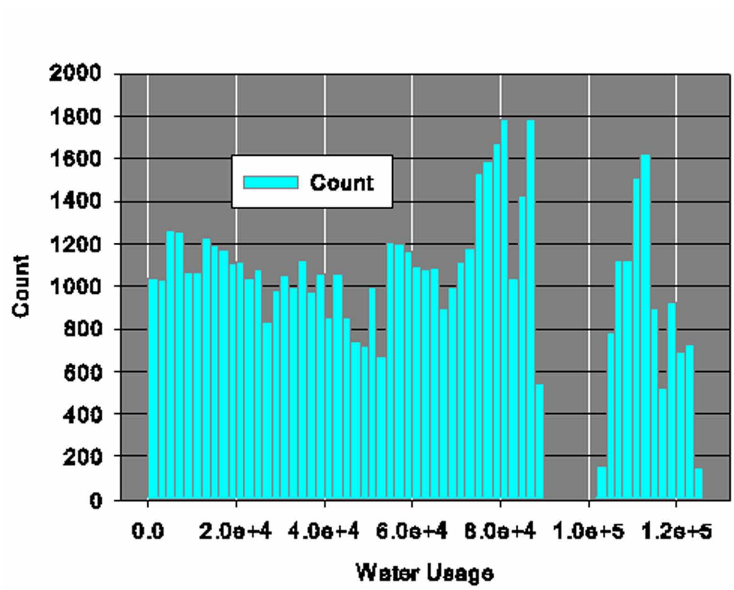
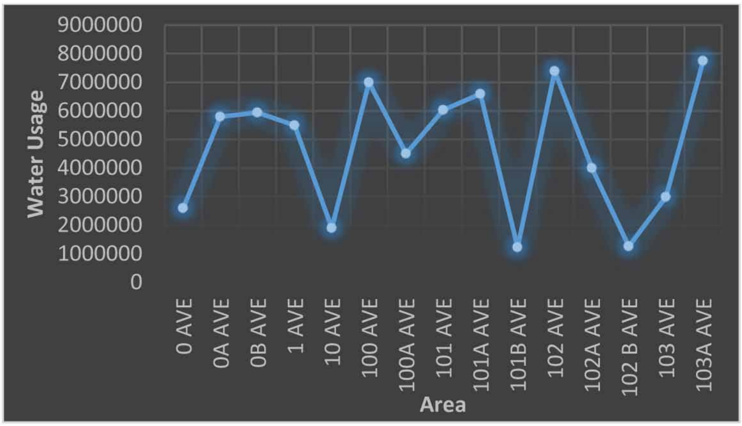


Figure 9. Water usage with respect to area



$$Sk = \sum_{i=1}^{N} \frac{1\left(Wi - Wn\right)3 \,/\, N}{s^3} \tag{1}$$

Where W is the mean, s is the standard deviation, and N is the number of data points (water consumption reading). While computing the skewness, the s is computed with N, rather than N-1.

Total 61263 houses are taken where average consumption for each house is 57877.937. However, the median analysis shows that 50% of the citizens consume less than 58186 cubic meter of water, 25% residents use less than 26893 while the 75% of the houses use less than 81983 cubic meter of water. The data is positively skewed, which means more than 50% houses' consumption is more than the average consumption. These analysis also helps the authorities to control the water resources based on their reservoirs. For instance, having more water reservoirs needs less storage of fresh water. Similarly, scarcity of water reservoir needs a proper mechanism to predict and store the required water for few days even. Thus, on the basis of average water consumption at each house, authorities can decide about the normal charges for each house that does not affect the common citizen. Similarly, we can manage energy such as electricity and gas as well by such analysis on them.

4.2.3. Flood Data Analysis

G. R. Brakenridge (2010) generated the complete flood dataset covering all the flood occurrences in the world, which contains information such as date of flood, area of flood, damage, intensity, death etc. The main causes of floods around the world are shown in Table 2. While analyzing flood data, authors observed that the major cause of floods all around the world is excessive rain at 35%, and the next major reason is snow melting at 1.5%. Therefore, they proposed to place rain and snow sensors to predict the floods prior to its occurrence. From rain data, authorities can predict the next year rain. Similarly, on maintaining snow data they can predict floods triggered when the snow is melt due to melting of snow. Moreover, the water reservoirs and the use of water can also be managed by measuring rain and snow.

4.2.4. Pollution analysis based on environmental data

The pollution dataset is generated by employing air quality gas sensors at the same places where the sensors for vehicular traffic measurement are placed at Aarhus city. The sensors are placed at 449

Table 2. World flood report from 1985-2014

Flood Type	Total Floods	Duration	Total Deaths	%age of total floods
Avalanche	3	11	33	0.005
Rain	3657	41637	190426	35.48
Snow	134	2404	851	1.54
Storm	83	981	6320	0.94
Dam Break	54	568	3600	0.32
Typhoon	5	38	1486	0.056

different locations, and the data is generated for the period. The levels of the following toxic gases in the air is measured: Nitrogen dioxide, Ozone, nitrogen oxide, carbon dioxide, and particle matters, etc.

Ozone (O3) is Amalgamation of three atoms of oxygen. It upsets the living tissues of humans when it interacts with them, such as, it can damage your lungs, cause a cough, an annoyed throat, result to a sunburn in the lungs etc. O3 is generated by the combination of Nitrogen oxide (NO), volatile organic compounds (VOC), and Nitrogen dioxide (NO2). So, the Nitrogen dioxide is also hazardous as more as VOC's and NO2 cause higher ozone level. Sulfur dioxide (SO2) raises asthma signs and also adversely affects the bronchoconstriction. "Particulate matter" is a multifaceted merging of liquid droplets and minor atoms. The particle can be made by acids, organic chemicals, or dust particles. These are so small that they can penetrate into the lungs and cause severe health complications.

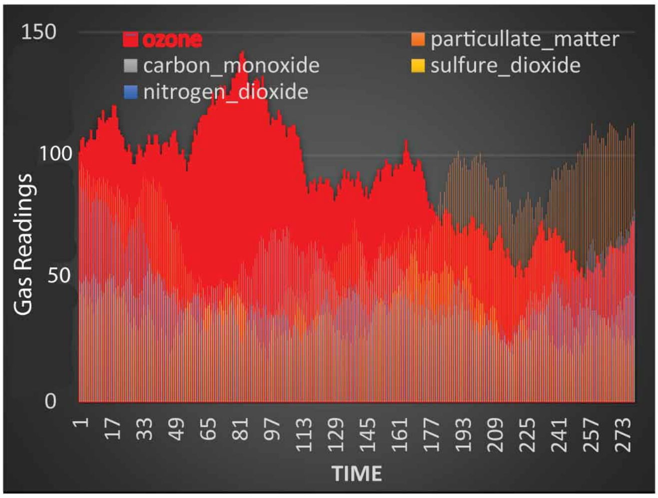
One of the transportation study (Eurostat, 2007) shows that each citizen travels at least one hour per day in Europe. This increases the number of transportation resources such as buses, trucks, trains, cars, motorbikes etc. The means of transportation is the foundation of 12% Co2 (Eurostat, 2011). Moreover, road pollution is two times more dangerous as car accidents (Yim & Barrett, 2012). In addition, the vehicular pollution in the air also spoils the youth health condition and raises the risk of prior deaths (US environmental Protection Agency (EPA), 2012). The main gasses in the air that affect the human health are ozone (O3), carbon monoxide, sulfur dioxide (SO2), nitrogen oxide, and particulate matter.

To keep the gasses value within a limit, the values of gas are little modified (Bischof et al., 2014; Tönjes et al., 2014; Kolozali et al., 2014). However, it will not affect the analysis, authenticity and outcome of the gasses. The values are calculated as:

- At the start a value between 25 and 100 is assigned. After every 5 minutes, the values are updated as follows:
- If the value were below 20, the new value is the last value + random integer between 1 and 10
- If the value were higher than 210, it would now be the last value - random integer between 1 and 10
- Else the value is updated as the last value + a random integer between -5 and 5.

These gasses are hazardous when their values are greater at peak level. Figure 10 shows the pollution level in Aarhus city. The ozone value at time 70 to 90 is very harmful, which is shown as red peak area. Similarly, particulate matters value at time 185-215 are also very severe for health at more than 245. Moreover, the Nitrogen dioxide at the start and the end of time interval might affect the children's health. The government authorities should do Super City planning by analyzing the past and changing behaviors of the pollution in different seasons and month. If there is more pollution recorded most of the time in the year at a particular location, then government should transfer its industries from that place, reduce traffic, and perform other necessary actions to reduce the pollutants level. They can shift industries outside the cities or build new industries at remote areas when these pollutants start growing.

Figure 10. Pollution level at different time of the day

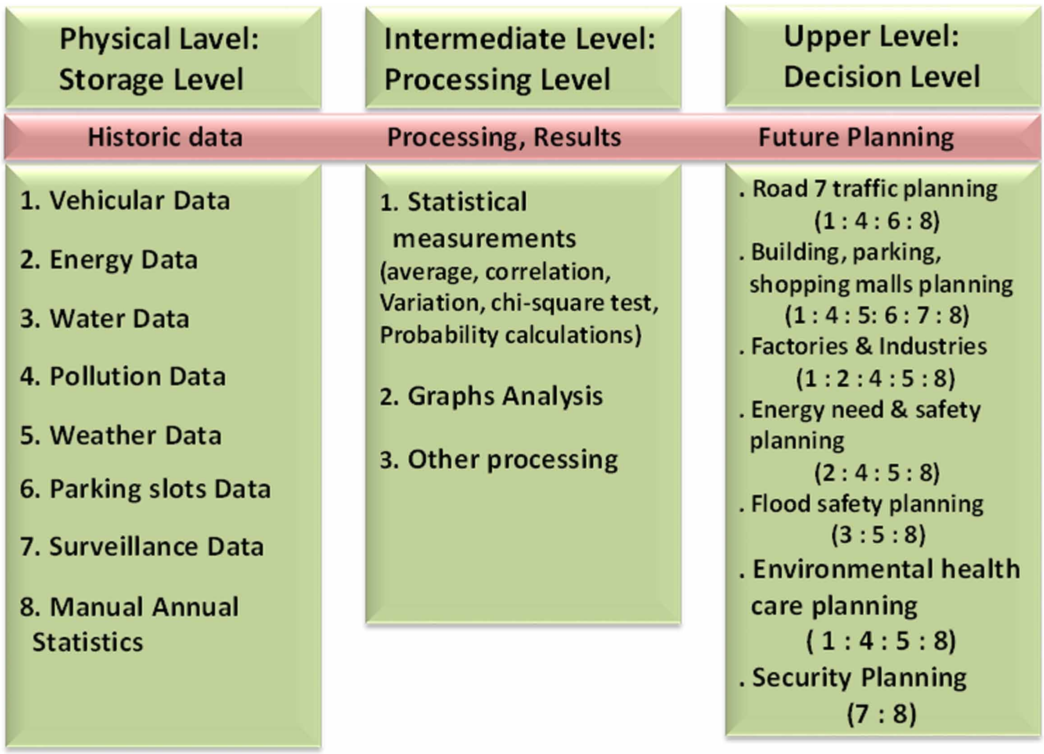


5. SYSTEM IMPLEMENTATION

In this section authors describe the implementation of the system, through which they used IoT-based Big Data for Super City planning. Hadoop ecosystem is used in order to efficiently handle and process Big Data. The system is implemented based on the analysis made on various datasets using the proposed system architecture, which is implemented using Hadoop single node at Ubuntu 14.04 LTS with 3.2 GHz x 4 processors and 4 GB memory. Since the network traffic in packet format is not efficiently processed by Hadoop. Therefore, Hadoop-pcap-lib, Hadoop-pcap-scr-de APIs are used to process the network traffic to perform real-time analysis for Smart City. The network traffic is then converted into sequence file to store on Hadoop Distributed File System (HDFS). All the historical information are stored on Hadoop HDFS for Super City planning. Here, authors are mainly interested in Super City planning rather than Smart city development, which gives importance to offline data analysis. The system is composed of various modules, which includes road traffic monitoring, building construction and road construction, factory and industries planning, energy planning and home safety, security planning, flood safety planning, health planning, etc.

The proposed Super City planning system is implemented in three level i.e. physical level, intermediate level, and upper level as shown in Figure 11. Physical level is the HDFS storage, which is the Big Data source for Super City planning. All the historic IoT data is stored in sequence file format at physical level after necessary action including filtration, classification, preprocessing is performed. In the figure, authors have given a number to each data source such as Number 1: vehicular data, number 2: energy data, and so on. The 2nd level is the intermediate level, which handles the data processing and computation of the data, such as statistical measures, graph processing etc., the Third level is the upper level which handles the decision-making operation. All machine learning, decision model algorithms, rule-based algorithm etc. are implemented here which make a decision on the results generated by the intermediate level. In short all Super City planning is done on this level based on the results generated by level two using physical level dataset. At this level, each planning decision that is made, such as road building, factory industries planning, security planning etc., have separate modules or functions. In addition, the number written under the module shows the dataset used by the module e.g., the road traffic management module takes the data from source 1, 4, 6, and 8, which are the vehicle data, pollution data, parking lots and annual statistics datasets, respectively.

Figure 11. Super city planning system implementation scenario



6. SYSTEM EVALUATION

The main focus of the system development is the processing of Big Datasets. Therefore, while implementing and evaluating the system, authors primarily focus on the efficiency of the system. MapReduce programming is used for performing analysis for Super City planning using single node Hadoop setup as described in section 5. They take various datasets to evaluate the efficiency of the system. Later to make the datasets more complex, they merged various heterogeneous datasets into a single dataset.

The data processing time (in ms) and the system throughput (in Mbps) are main parameters used for measuring the efficiency of the system. Figure 12 shows the processing time of various datasets corresponding to the dataset size. The throughput results are depicted in Figure 13. The processing time is proportional to the data size. However, the system throughput is inversely proportional to the volume of the processed data. The inversely proportional throughput is resulted because of the parallel processing environment of the Hadoop ecosystem and the powerful filtration process, which is the main accomplishment of the proposed system. Moreover, anyone can look at the processing of largest dataset i.e. 5345MB, where the data processing takes less than 300000ms time, which is quite high.

7. CONCLUSION

City planning has a key influence on the development of the nations which increases the decision control of the societies by making a smart decision. In this paper, authors coined new term for

Figure 12. Data Processing time on increasing-size vehicular datasets

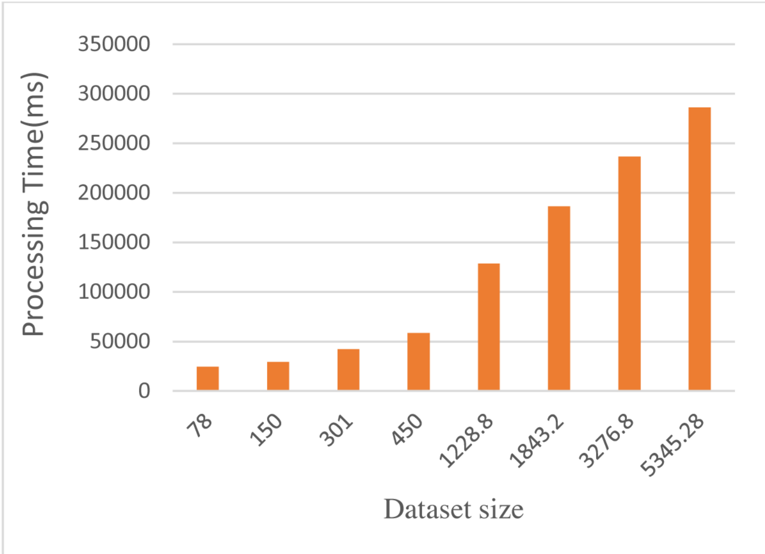
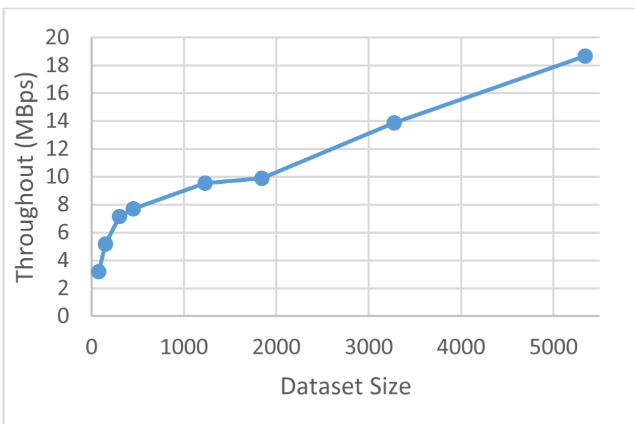


Figure 13. Throughput of the system on various increasing-size datasets



intelligent, smart and ideal city as Super City. Later, they proposed a complete system for Super City planning that uses the services from other smart systems using IoT. The smart systems generated Big Data is processed and analyzed by the proposed implementation model using Hadoop. An architecture with four tiers is offered that devises the functionality of data aggregation and collection, data communication, data processing and management, and data interpretation. Various datasets generated by different smart systems are used for urban data analysis, data investigation, and system evaluation. The proposed system would have a major impact on the life of the coming generations of citizens by providing many facilities and needs prior to their demand. The system is evaluated based on the performance efficiency while focusing on the data processing time and the system throughput. The designed system performs extraordinarily when it process larger datasets. The throughput is rises as the data size increases. In future, authors are planning to extend their work by providing practical integration of the smart systems with the main municipality's system to achieve Super City. They planned to start from vehicular network integrations and then proceed to smart home integration. As the system is highly critical, they are also interested in providing security to the system.

ACKNOWLEDGMENT

This research was supported by Kyungpook National University Bokhyeon Research Fund, 2015.

REFERENCES

Ahmad, A., Paul, A., Rathore, M. M., & Chang, H. (2016). Smart cyber society: Integration of capillary devices with high usability based on Cyber–Physical System. *Future Generation Computer Systems*, *56*, 493–503. doi:10.1016/j.future.2015.08.004

Bischof, S., Karapantelakis, A., Nechifor, C. S., Sheth, A. P., Mileo, A., & Barnaghi, P. (2014). Semantic modelling of smart city data. *Position Paper in W3C Workshop on the Web of Things: Enablers and services for an open Web of Devic*es, Berlin, Germany (pp 25-26).

Brakenridge, G. R. (2010). Global Active Archive of Large Flood Events. *Dartmouth Flood Observatory, University of Colorado*. Retrieved from http://floodobservatory.colorado.edu/Archives/index.html

Bulkeley, H., McGuirk, P. M., & Dowling, R. (2016). Making a smart city for the smart grid? The urban material politics of actualising smart electricity networks. *Environment and Planning*.

Dixit, S., & Prasad, R. (Eds.). (2007). *Technologies for home networking*. John Wiley & Sons.

Dodge, M., & Kitchin, R. (2007). The automatic management of drivers and driving spaces. *Geoforum*, *38*(2), 264–275. doi:10.1016/j.geoforum.2006.08.004

Eurostat. (2007). Passenger mobility in Europe. *European Commission*.

Eurostat. (2011). Energy, transport and environment indicators. European Commission.

Evans, D. (2011, July). The Internet of Things. *Cisco Blog*. Retrieved from http://blogs.cisco.com/news/the-internet-of-things-infographic

Giroux, S., & Pigot, H. (2005). From Smart Homes to Smart Care: ICOST 2005. *Proceedings of the3rd International Conference on Smart Homes and Health Telematics* (Vol. 15). IOS Press.

Gramaglia, M., Trullols-Cruces, O., Naboulsi, D., Fiore, M., & Calderon, M. (2014, June). Vehicular networks on two Madrid highways. *Proceedings of the2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON),* Singapore (pp. 423-431). doi:10.1109/SAHCN.2014.6990380

Han, S. S. (2005). Global city making in Singapore: A real estate perspective. *Progress in Planning*, *64*(2), 69–175. doi:10.1016/j.progress.2005.01.001

Hancke, G. P., & Hancke, G. P. Jr. (2012). The role of advanced sensing in smart cities. *Sensors (Basel, Switzerland)*, *13*(1), 393–425. doi:10.3390/s130100393 PMID:23271603

ITU-T Focus Group on Smart Sustainable Cities (2015, March). Standardization roadmap for smart sustainable cities. Telecommunication Standardization Sector of ITU.

Jin, J., Gubbi, J., Marusic, S., & Palaniswami, M. (2014). An information framework for creating a smart city through internet of things. *IEEE Internet of Things Journal*, *1*(2), 112–121. doi:10.1109/JIOT.2013.2296516

Kolozali, S., Bermudez-Edo, M., Puschmann, D., Ganz, F., & Barnaghi, P. (2014, September). A knowledge-based approach for real-time IoT data stream annotation and processing. *Proceedings of the2014 IEEE International Conference on Internet of Things (iThings), Green Computing and Communications (GreenCom), and Cyber Physical and Social Computing (CPSCom),* Taipei, Taiwan (pp. 215-222). doi:10.1109/iThings.2014.39

Kyriazis, D., Varvarigou, T., White, D., Rossi, A., & Cooper, J. (2013, June). Sustainable smart city IoT applications: Heat and electricity management & Eco-conscious cruise control for public transportation. *Proceedings of the2013 IEEE 14th International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)* (pp. 1-5).

Naboulsi, D., & Fiore, M. (2013, July). On the instantaneous topology of a large-scale urban vehicular network: the cologne case.*Proceedings of the fourteenth ACM international symposium on Mobile ad hoc networking and computing*, Bangalore, India (pp. 167-176). doi:10.1145/2491288.2491312

Odroma, M., & Ganchev, I. (2010). The creation of a ubiquitous consumer wireless world through strategic ITU-T standardization. *IEEE Communications Magazine*, *48*(10), 158–165. doi:10.1109/MCOM.2010.5594691

Rathore, M. M., Ahmad, A., Paul, A., & Thikshaja, U. K. (2016, May). Exploiting real-time big data to empower smart transportation using big graphs. *Proceedings of the2016 IEEE Region 10 Symposium (TENSYMP),* Bali, Indonesia (pp. 135-139). doi:10.1109/TENCONSpring.2016.7519392

Section, G. I. S. (2014, February). Water Meters. *Surrey's Open Data Program repository*. Retrieved from http://data.surrey.ca/dataset/water-meters

Singer, N. (2012). Mission control, built for cities: IBM Takes 'Smarter Cities' Concept to Rio de Janeiro. *New York Times*.

Srivastava, L. (2004). Japan's ubiquitous mobile information society. *Info, 6*(4), 234-251.

Tönjes, R., Barnaghi, P., Ali, M., Mileo, A., Hauswirth, M., Ganz, F., . . . Puiu, D. (2014). Real time iot stream processing and large-scale data analytics for smart city applications. *Proceedings of theEuropean Conference on Networks and Communications (*poster session), Bologna, Italy.

Uppoor, S., & Fiore, M. (2011, November). Large-scale urban vehicular mobility for networking research. Proceedings of the2011 IEEE Vehicular Networking Conference (VNC), Amsterdam, Netherlands (pp. 62-69). doi:10.1109/VNC.2011.6117125

Uppoor, S., Trullols-Cruces, O., Fiore, M., & Barcelo-Ordinas, J. M. (2014). Generation and analysis of a large-scale urban vehicular mobility dataset. *IEEE Transactions on Mobile Computing*, *13*(5), 1061–1075. doi:10.1109/TMC.2013.27

US environmental Protection Agency (EPA) (2012). How Air Pollution Affects the View. Retrieved from https://www.epa.gov/sites/production/files/2015-05/documents/haze_brochure_20060426.pdf

Xia, F., Yang, L. T., Wang, L., & Vinel, A. (2012). Internet of things. *International Journal of Communication Systems*, *25*(9), 1101–1102. doi:10.1002/dac.2417

Yim, S. H., & Barrett, S. R. (2012). Public health impacts of combustion emissions in the United Kingdom. *Environmental Science & Technology*, *46*(8), 4291–4296. doi:10.1021/es2040416 PMID:22435408

Zeng, D., Guo, S., & Cheng, Z. (2011). The web of things: A survey. *Journal of Communication*, *6*(6), 424–438.

Muhammad Mazhar Ullah Rathore received his Master's degree in Computer and Communication Security from the National University of Sciences and Technology, Pakistan in 2012. Currently, he is pursuing his PhD with Dr. Anand Paul at Kyungpook National University, Daegu, South Korea. His research includes Big Data Analytics, Network Traffic Analysis and Monitoring, Remote Sensing, IoT, Smart City, Urban Planning, Intrusion Detection, and Computer and Network Security. He is an IEEE student member. He is also a nominee of Best Project Award in 2015 IEEE Communications Society Student Competition for his project "IoT based Smart City". He is serving as a reviewer for various IEEE, ACM, Springer, and Elsevier journals.

*Anand Paul, (SM'**) received the PhD degree in Electrical Engineering from the National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently working as an Associate Professor in the School of Computer Science and Engineering, Kyungpook National University, South Korea. He is a delegate representing South Korea for M2M focus group and for MPEG. His research interests include Algorithm and Architecture Reconfigurable Embedded Computing. He is IEEE Senior member and has guest edited various international journals and he is also part of Editorial Team for Journal of Platform Technology, ACM Applied Computing review and Cyber–Physical Systems. He serves as a reviewer for various IEEE /IET/Springer and Elsevier journals. He is the track chair for Smart human computer interaction in ACM SAC 2015, 2014. He was the recipient of the Outstanding International Student Scholarship award in 2004–2010, the Best Paper Award in National Computer Symposium and in 2009, and International Conference on Softcomputing and Network Security, India in 2015.*

Awais Ahmad received his BS (CS) from the University of Peshawar and Masters (Telecommunication and Networking) from Bahria University, Islamabad Pakistan in 2008 and 2010 respectively. During his Master's research work he worked on energy efficient congestion control schemes in Mobile Wireless Sensor Networks (WSN). Currently, he is pursuing his PhD at Kyungpook National University, Daegu South Korea. There he got research experience on Big Data Analytics, IoT/SIoT, 4G/5G, Machine-to- Machine Communication, and Wireless Sensor Network. He also received four prestigious awards: (1) IEEE Best Research Paper Award: International Workshop on Ubiquitous Sensor Systems (UWSS 2015), in conjunction with the Smart World Congress (SWC 2015), Beijing, China, (2) Research Award from President of Bahria University Islamabad, Pakistan in 2011, (3) Best paper nomination award in WCECS 2011 at UCLA, USA, and (iv) Best paper award in the 1st Symposium on CS&E, Moju Resort, Korea in 2013

Gwanggil Jeon received the PhD degree in Department of Electronics and Computer Engineering from Hanyang University, Seoul, Korea, in 2008. From 2008 to 2009, he was with the Department of Electronics and Computer Engineering, Hanyang University, from 2009 to 2011, he was with the School of Information Technology and Engineering (SITE), University of Ottawa, as a postdoctoral fellow, and from 2011 to 2012, he was with the Graduate School of Science & Technology, Niigata University, as an associate professor. He is currently an Associate Professor with the Department of Embedded Systems Engineering, Incheon National University, Incheon, Korea. His research interests fall under the umbrella of image processing, particularly image compression, motion estimation, demosaicking, and image enhancement as well as computational intelligence such as fuzzy and rough sets theories.