When Big Data meets Smart Cities

Amjad Mahfoud
amjadoof@gmail.com (Faculty of Informatics Engineering, Damascus University, Damascus Syria)

Abstract:

The rapid growth in the population density in urban cities demands that services and infrastructure be provided to meet the needs of inhabitants. Thus, there has been an increase in the request for embedded devices, such as sensors, smartphones, etc. leading to an incredible amount of generated data to benefit from. Managing this amount of data could lead to improving the health field, the traffic and the public services in the modern cities, this paper is meant to address this subjects and find the proper way to make this data useable and helpful in designing and planning modern 'Smart' cities.

Keywords: Big Data, Internet of Things, Cloud Computing, Smart Cities, Analysis, planning.

I. Introduction

THE significant increase in connected devices and sensors as well as the increasingly raising computational performance have made the vision for living in a smart environment feasible. Several applications of a smart environment have been introduced recently, including smart homes [1], smart grids [2], smart transportation [3], smart healthcare [4], and smart cities [5], smart urban performance doesn't only depend on physical infrastructure but also on the availability and the quality of communication infrastructure and the data transported via the network. The key enabler of these smart city applications is possibly the IoT in which everyday objects and devices are connected. This enormous generated amount of data is the core of the services rendered by the IoT.

At present, a large amount of data is being generated from different data sources, such as smartphones, computers, sensors, cameras, global positioning systems, social networking sites, commercial transactions, and games. Given that the data generated in our present digitized world continuously grow, efficient data storage and processing facilities have posed challenges to the traditional data mining and analytics platforms.

Big data analytics can extract meaningful information from the oceans of data produced by sensor devices. Effective analysis and utilization of big data is a key factor for the success in many business and service domains, including the smart city application. The application of big data in a smart city has many benefits and challenges, including the availability of large computational and storage facilities to process streams of data produced within a smart city environment. One of the possible means to tap this benefit is a reliance on cloud computing services and IoT technologies.

Big data offers the potential for the city to obtain valuable insights from a considerable amount of data collected through various sources. Certainly, the characteristics of such data mostly include unstructured features compared with big data collected by other means [6].

Figure 1 illustrates the landscape of the smart technologies with big data and cloud computing, in which various smart applications exchange information using embedded sensor devices, etc… that generate large amounts of unstructured data. These large amounts of unstructured data are collected and stored in the cloud or data center using distributed fault tolerant databases such as NO SQL or Systems like Hadoop and Spark that have the ability to maintain and manipulate such huge amount of data for further use and analysis

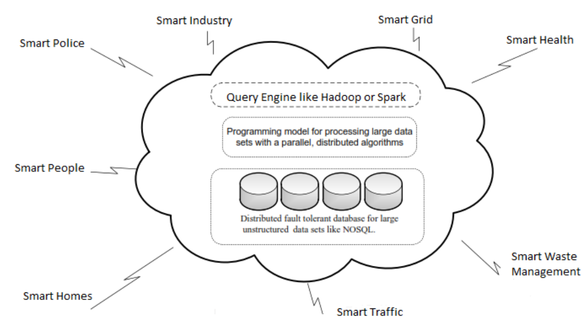

Fig. 1 Landscape of the smart city and big data technologies

II. State of the Art Technologies

A. The Internet of Things

The IoT has witnessed recent implementation in the smart city to develop intelligent systems such as smart grids, smart retail, smart homes, smart water, smart transportation, smart healthcare, and smart energy.

However, a universally agreed definition of a smart city is yet to be conceived, and recognizing common global trends is challenging. The smart city focuses on applying the next-generation information technology to all walks of life, embedding sensors and equipment in hospitals, power grids, railways, bridges, tunnels, roads, buildings, water systems, dams, oil and gas pipelines as well as other objects worldwide, thereby forming the IoT [8].

The Internet revolution led to the interconnection between people at an unprecedented scale and pace. The next revolution will be the interconnection between objects to create a smart city. The smart city emphasizes the interconnection of sensing and actuating devices, thereby allowing information sharing across platforms through a unified framework. Such sharing is achieved by seamless ubiquitous sensing, data analytics, and information representation with cloud computing as the unifying framework.

The present is the post-PC era, in which smartphones and other handheld devices are changing our environment by making it increasingly interactive and informative [8].

B. Big Data

Big Data systems are stored, processed, and mined in smart cities efficiently to produce information to enhance different smart city services. Big data can achieve its goals and can advance the services in smart cities using the right tools and methods for efficient and effective data analysis. Such effectiveness will encourage collaboration and communication between entities and can facilitate the creation of additional services and applications that can further enhance the smart city [8].

C. Cloud Computing

Cloud computing is used to describe a variety of different types of computing models that involve many computers or clusters connected through a real-time communication network. Cloud computing provides services to perform complex large-scale computing tasks such as mining big social network data generated through smartphone. Cloud computing services, such as platform as a service, software as a service, and infrastructure as a service, can be combined with IoT. Cloud computing also provides the underlying engine via the big data technology such as Hadoop framework [8] [12].

D. Hadoop

Hadoop was introduced to provide an enabling platform and programming models for the distributed processing of large datasets across different clusters. Hadoop comprises two primary components: Hadoop Distributed File System (HDFS) and MapReduce, which are closely related to each other [7]. Although the real-time requirements of data storage and processing in the smart city are considered, the adoption of streaming architecture will guarantee the efficient and seamless communication between sensing devices within the smart city network. Such technology has been adopted recently with the introduction of many stream processing platforms, such as Apache S4, Storm, and Spark streaming, which can enable data storage and processing across various interconnected nodes [8].

E. Spark

The main abstraction in Spark is that of a resilient distributed dataset (RDD), which represents a read-only collection of objects partitioned across a set of machines that can be rebuilt if a partition is lost. Users can explicitly cache in RDD in memory across machines and reuse it in multiple MapReduce-like parallel operations. RDDs achieve fault tolerance through a notion of lineage: if a partition of an RDD is lost, the RDD has enough information about how it was derived from other RDDs to be able to rebuild just that partition. Although RDDs are not a general shared memory abstraction, they represent a sweet-spot between expressivity on the one hand and scalability and reliability on the other hand, and were found to be well-suited for a variety of applications. [9].

F. Batch Processing

In order to cope with the volume challenge so-called noSQL databases are gaining in popularity. Those databases can be classified according to their scalability with respect to data size and suitability to model complex data relationships [10].

With decreasing ability to handle data size and increasing capability to handle complex relationships, these stores can be distinguish: key-value stores, columnar stores, document databases, and graph databases. Key-value stores are databases that scale easily, but are only able to capture simple relationships, i.e., a key and its value. Columnar stores in contrast store data in columns and are thus better suited for queries that access only small portions of high dimensional data. Document databases are able to store even more complex data, but typically do not scale as well. Graph databases in contrast can easily model any relationship, but do not scale as easily [10].

The actual processing of data uses different computational frameworks such as Hadoop’s MapReduce, Hama’s Bulk Synchronous Processing framework, or graph processing frameworks. They are designed to process large volumes of data in batches, i.e., in regular intervals.

As a consequence these technologies are not suitable to process data in real-time [10].

III. Applications of Big Data in Smart Cities

The application of big data technologies for the smart city enables efficient data storage and processing to produce information that can enhance different smart city services. In addition, big data helps decision makers plan for any expansion in smart city services and resources. For big data to achieve its goals and advance services in smart cities, it needs the right tools and methods for efficient and effective data analysis. These tools and methods may encourage collaboration and communication between entities and provide services to many sectors in the smart city, as well as improve customers’ experiences and business opportunities [8].

A. Smart Grid.

The rapid distribution of smart grids has enabled researchers to integrate, analyze, and use real-time power generation and consumption data, as well as other types of environmental data. The improvement in energy efficiency and intelligent services is expected to result in a high investment efficiency of the existing smart grid infrastructure. In a smart grid environment, a large amount of data is generated from different sources, such as the power utilization habits of users, the efficient use of big data collected from the smart grid environment can help decision makers come up with a wise decision in terms of the supply level of electricity while fulfilling the demands of the user [8] [14].

B. Smart Healthcare.

In the past decade, an enormous amount of data has been generated in the healthcare sector. The rapid rate of increase in the world's population has facilitated the rapid changes in the models of treatment delivery, and many decisions behind those changes are driven by data. Proper analytics tools can allow healthcare specialists to collect and analyze patients' data, which can likewise be used by insurance agencies and administration organizations. Moreover, proper analytics of big healthcare data can help predict epidemics, cures, and diseases.

Figure 2 Illustrates the Smart Hospital System (SHS) the system is composed of three main parts: (1) the RFID-enhanced wireless sensor network, named Hybrid Sensing Network (HSN) hereafter, (2) the IoT Smart Gateway, and (3) the user interfaces for data visualization and management [15].

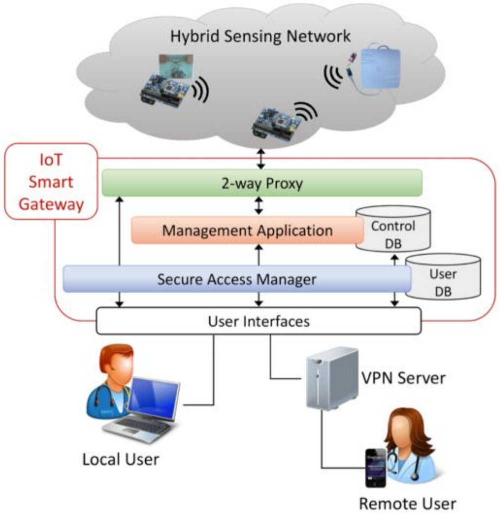

Fig. 2 Overview of the Smart Hospital System (SHS) architecture [15].

C. Smart Transportation.

The patterns obtained from the large amounts of traffic data can help improve transportation systems in terms of minimizing traffic congestion by providing alternative routes and reducing the number of accidents by analyzing the history of mishaps, including factors such as their cause and the driver speed. The data generated by transport systems can also help optimize freight movements. Moreover, the big data collected from smart transport systems can help consolidate shipments and optimize shipping movements by reducing supply chain wastage. Smart transport data can also provide many benefits, such as reducing the environmental impact and increase safety as well as improving end-to-end user experience, among many others [8] [16].

IV. Challenges

The challenges are split into two main categories Big Data-related and security-related.

A. Big Data Related

Big Data challenges and technologies can best be described along the so-called 3 V's: Volume, Velocity, and Variety [10].

The Volume challenge: Volume refers to storing, processing, and quickly accessing large amounts of data. While it is hard to quantify the boundary fora volume challenge, common data sets in the order of hundreds of Terabyte or more are considered to be big.

The Velocity Challenge: Velocity refers to the fact that data is streaming into the data infrastructure of the enterprise at a high rate and must be processed with minimal latency.

The Variety Challenge: In a data-driven economy the objective is to maximize the business value by considering all available data. From a technical perspective one could formulate that problem by evaluating a function over all accessible data sets

B. Security Related

Buggy, brittle and hackable cities: The embedding and use of ubiquitous and pervasive computing in city environments is creating city services and spaces that are dependent on software to function. Where in software and the spatiality of everyday life become mutually constituted, so that if the software fails a space is not produced as intended as the old analogue system and associated tacit knowledge has been entirely replaced. For example, if the software used to control a subway system crashes, then the trains do not run happened in many cities in the past few years [12].

V. Proposed Structure Of Big Data in Smart Cities

With smart city applications producing continuous large data from heterogeneous sources, existing relational database technologies are inadequate to handle such huge amounts of data given the limited processing speed and the significant storage expansion cost. To address this problem, big data processing technologies, which are based on distributed data management and parallel processing, have provided enabling platforms for data repositories, distributed processing, and interactive data visualization [8].

The system structure of big data in the smart city, as shown in Figure 2, can be divided into multiple layers to enable the development of integrated big data management and smart city technologies. Each layer represents the potential functionality of big data smart city components. The first layer is the set of objects and devices connected via local and/or wide-area networks. Most of these objects and devices actively generate a huge amount of unstructured data every second. In the second layer, all the collected unstructured data are stored in a shared distributed fault-tolerant databases located either in the city data center equipped with all network elements or by big data storage such as S3 [8],

In stream processing, data must be processed quickly so that companies and individuals can react to changes in real time in a smart city environment. Many technologies can help process and act on real-time streaming unstructured data in real time such as Spark, Storm, and S4 [8].

The smart analysis shown in Figure 1 can be designed using scalable machine learning algorithms or other novel data mining algorithms to provide extraction of patterns and knowledge from large amounts of data. A Typical example of such technology is Apache Mahout, in which many machine learning libraries for data filtering, clustering, and classification can be found. The last layer is the application services, in which people and machines directly interact with each other to make smart decisions. Such applications can be used for different purposes such as recommendation, fraud detection, sentiment analysis, intelligent traffic management, and web display analysis [8].

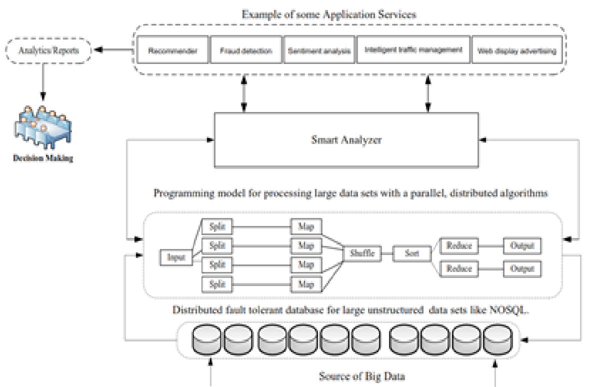

Fig. 3 Construction frame of big data technologies in smart city

VI. Construction of a Smart City

Smart City will be the future trend of urban development. Generally, the construction of smart city can be divided into three levels, including the construction of public infrastructure, construction of public platform for smart city, the construction of application systems. In this three-level, the construction of application systems is particularly important, and has earned great concern across the country. Currently, in addition to defense and national security applications, smart city has been typically applied to various aspects. Figure.3 shows some typical applications [11].

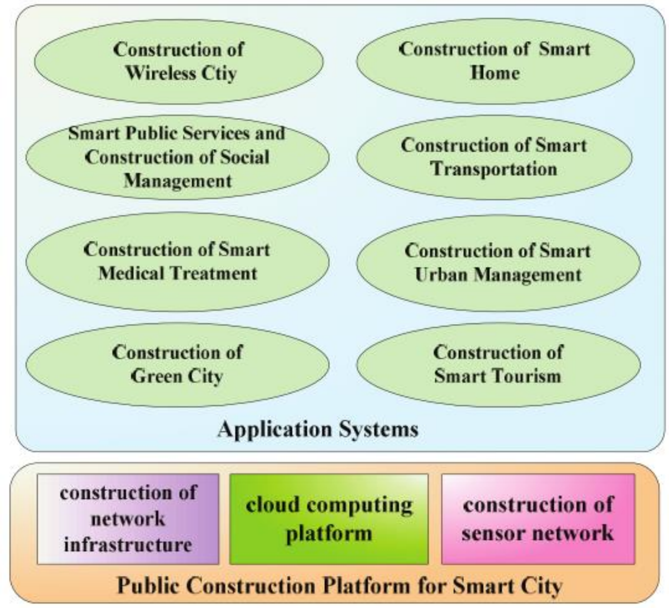

Fig. 3 Construction Frame of Application Systems for Smart City [11]

References

[1] Caragliu, A., Del Bo, C., & Nijkamp, P. (2011). Smart cities in Europe. *Journal of urban technology, 18*(2), 65-82.

[2] Chen, S.-y., Song, S.-f., Li, L., & Shen, J. (2009). Survey on smart grid technology. Power System Technology, 33(8), 1-7

[3] Adeli, H., & Jiang, X. (2009). Intelligent infrastructure: neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures: Crc Press.

[4] Demirkan, H. (2013). A smart healthcare systems framework. *It Professional, 15*(5), 38-45. (Caragliu et al., 2011; Chourabi et al., 2012)

[5] Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K Scholl, H. J. (2012). *Understanding smart cities: An integrative framework.* Paper presented at the System Science (HICSS), 2012 45th Hawaii International Conference on.

[6] Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171-209

[7] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: review and open research issues. *Information Systems, 47*, 98-115. doi: http://dx.doi.org/10.1016/j.is.2014.07.006

[8] Hashem, Ibrahim I. A. T., et al. "The Role of Big Data in Smart City." International Journal of Information Management 36.5 (2016): 748-758

[9] Zaharia, Matei, et al. "Spark: Cluster computing with working sets." *HotCloud*10.10-10 (2010): 95.

[10] Strohbach, Martin, et al. "Towards a Big Data Analytics Framework for IoT and Smart City Applications." Modeling and processing for next generation big data technologies. Springer International Publishing 2015. 257-282.

[11] Kehua Su, Jie Li, Hongbo Fu. "Smart City and the Applications." Electronics, Communications and Control (ICECC), 2011 International Conference on, IEEE, 2011.

[12] Kitchin, Rob. "The real-time City? Big data and smart urbanism." GeoJournal 79.1 (2014): 1014.

[13] Schaffers, Hans, et al. "Smart cities and the future internet: Towards cooperation frameworks for open innovation." *The future internet* (2011): 431-446.

[14] Amin, S. Massoud, and Bruce F. Wollenberg. "Toward a smart grid: power delivery for the 21st century." *IEEE power and energy magazine* 3.5 (2005): 34-41.

[15] Catarinucci, Luca, et al. "An IoT-aware architecture for smart healthcare systems." *IEEE Internet of Things Journal* 2.6 (2015): 515-526.

[16] Wang, Yun, et al. “A Big Data Approach for Smart Transportation Management on Bus Network.” Smart City Conference (ISC2), 2016 IEEE International. IEEE, 2016.