Management of Big Geographic Data for Smart City Applications

Arif Cagdas AYDINOGLU and Rabia BOVKIR, Turkey

Key words: Smart city, Big Data, Data Modelling, Geographic Data Management,

SUMMARY

It is supposed that more than the half of the world's population live in urban areas today and it is estimated that this number will reach 80% by 2050. Significant problems such as traffic, transportation, air and environmental pollution, emergency response, immovable management etc. arise as a result of this intensity in urban population. For the solution of these problems "smart city" concept has emerged with the purpose of providing better service to the citizens by using all urban resources in an efficient and sustainable manner. In this paper, firstly the smart city concept is explained with all the requirements and application domains. Then the concept of spatially enabled big data and its importance for smart cities are emphasized. Secondly, initiatives to management and applications of big geographic data for smart cities will be introduced and explained. In order to use large volume of geographic data coming from different sources, big data technologies for enabling better processing and management for smart city applications will be analyzed. Big data solutions such as Hadoop ecosystem, database structures such as NoSQL and NewSQL and data analytics will be defined. Finally a general overview is given for the big geographic data solutions in smart city concept with example applications. In this study, it is aimed to give a literature review of the smart city and big data concepts to help the public authorities for determining the right investment plan for successful smart city implementation with high efficiency.

Management of Big Geographic Data for Smart City Applications

Arif Cagdas AYDINOGLU and Rabia BOVKIR, Turkey

1. INTRODUCTION

Today, more than 50% of the world's population live in urban areas, and the reports published by the United Nations (URL 1) reveal that this number will reach 80% by 2050. As a result of this concentration, the urban environment has become quite complicated. Especially in mega cities like Istanbul, significant problems concerning traffic and transportation, energy, water and waste management, air and environmental pollution, emergency response, and immovable management arise due to excessive population density. In this context, new ideas and debates about technology-based urban life and planning solutions in heavily urbanized areas have started as a solution to all these problems (Neirotti et al., 2013; Alawadhi et al., 2012; Nam and Pardo, 2011). The most disputed and accepted concept in these discussions was "Smart City" (Neirotti et al., 2013; Hollands, 2008). While there is not a standard and precise definition of this concept, in the most general sense it can be expressed as the modernization effort aiming to provide better service to the city-dwellers by using all urban resources in an efficient and sustainable manner. In other words, it is the way of monitoring and managing cities as a whole smarter system with the help of information technologies (Terzi and Ocakci, 2017).

Within the scope of smart cities, the ability of processing large amount of data has become very important for successful applications with the developments in information technologies. Today, producing knowledge has come to the forefront with the development and progress of technology and the internet. Regardless of the size of the storage area, data is generated at an ever-increasing rate. This rapid growth of data volume can be caused by the growing volume of social networking interactions, the increasing of the location-sensitive devices and "smart sensors" that capture and transmit information about the physical world (Chen et al., 2014). For example, only one of the social networking sites can generate tens of TB data every day or some institutions can process tens of TB data in every hour each day (Sagiroglu and Sinanc, 2013). In addition, video and media resources can also be added to these data. However, most of these phenomena on the internet are not always meaningful and processable and they are often regarded as 'information garbage'.

The concept called "Big Data" is based on the principle of extracting meaningful data from the data collection which is actually considered as information repository. The simplest definition of the big data term is identified by George et al. (2014) as “Physical tracks that people leave behind during the internet, smartphone, and social media account interactions they use every day and the massive amount of data that all these interactions are brought together.” Big data management within the smart city concept enables the collection of valuable information from wide variety of sources and is the ability to process such big data performing successful applications (Hashem et al., 2015). With smartphones that are entered our lives with today's developing technology, each individual can be considered as a sensor. So that instant content can be obtained from social networking sites with status notifications that citizens share. All these developments bring out another important subject in the scope of smart cities: "Internet of Things (IoT)". IoT can be defined as the infrastructure based on interoperable information and communication technologies which makes developed services possible by linking physical and virtual things/objects (Internet of Things Global Standards Initiative, 2012).

In this paper, firstly the smart city and spatially enabled big data concepts are explained with the requirements and application domains. Secondly, initiatives and applications of big geographic data are introduced and explained for smart cities. In order to use large volume of geographic data coming from different sources, big data solutions for enabling better smart city management are analysed. Big data solutions and data analytics are defined. Finally, a general overview is given for the big geographic data solutions in smart city concept with example applications. In this study, it is aimed to give a literature review of the smart city and big data concepts to help the public authorities for determining smart city strategy.

2. SMART CITY CONCEPT

It is aimed to explain and clarify the meaning of the smart city concept by examining its background, characteristics, key attributes, application domains and subdomains in this section.

2.1 Background and Application Domains of Smart Cities

A smart city enables efficient integration of physical, digital and human system for providing a sustainable, comfortable and inclusionary future for its citizens. Smart city concept has been launched for emphasizing the necessity and importance of Information and Communication Technologies (ICTs) especially in the last two decades (Schaffers, 2012). Feasibility, progress and flexibility on geographic location are provided with the development of ICTs which promoted to initiatives to build smart cities (Tao, 2015). ICTs can be used for connecting the cities by computer, cloud and transport networks to dissolve critical urban issues in healthcare, traffic, energy, accessibility, mobility, security etc.

The smart city concept includes solutions about optimizing the use and management of both tangible assets like transport networks, natural resources, energy distribution networks and intangible assets like intellectual capital in business sector, organizational capital of public agencies. In smart city literature there are mainly two different approaches based on two different factors (Neirotti et al., 2014). First approach supports the way cities govern themselves to optimize the domains that are more sensitive for a smarter use of resources. The main significance of this approach is on effective management of energy and transportation networks, waste management, pollution control and information processing in these fields with the help of ICTs (Neitrotti et al., 2014). Other approach is based more on bottom-up design in which cities enable the citizens to access data and make their own decisions. In this approach, ICTs plays a restricted role in providing sustainability and dealing operations because stressing the importance in “soft” domains related to prosperity, social inclusion policies, culture and education (Neitrotti et al., 2014).

According to the literature, urban living domains can be classified as “hard” and “soft” domains where ICTs play an important role as key enabling component. Hard domains refer to buildings both used as office or residential, energy and water management, natural resources, environment, transport networks, waste management, healthcare, public security, mobility and logistics; soft domains refer to culture, educational, economy, social inclusion and welfare, public administration and e-government (Washburn et al., 2010; Correia and Wünstel, 2011; Nam and Pardo, 2011; Chourabi et al., 2012). Hard domains provide the ability of a city to sense and act more applicable applications due to the use of wireless technologies and smart sensors to deal with the “big data” (McKinsey Global Institute, 2011; McAfee and Brynjolfsson, 2012). On the other hands, in soft domain areas, ICT has a restricted role and is not mainly aspired to process and unify real-time data.

2.2 Smart City Architecture and Components

In smart city concept, the main purpose is to provide more quality and prospering life conditions appreciated by its citizens through ICTs to enhance the utility of both urban and social services. The smart city concept involves a focus on ICTs and almost in every aspect of urban and daily life takes advantages of it. According to Boyd Cohen (USDN, 2014), smart city mainly consists of six parts: smart people, smart economy, smart mobility, smart living, smart environment and smart government. However, these parts can vary because of different application needs. Regardless of the application area, the smart city concept includes technology, data process and people. Considering the technology perspective, smart city involves four components (IDB, 2016):

- **Communication Interfaces** includes web services, portals, applications for sending and receiving information from people and providers via open platforms. This component promotes the citizen participation via discussion forums, mobile applications, thematic social networks and public structure transparency.
- **Integrated Operation and Control Centers (IOCC)** includes technological (computers, application monitors etc.), physical (operating and management rooms etc.) and process infrastructures, government agency personnel and service providers for solving the smart city issues. These components involve receiving, processing and analysing the sensor data and distribute integrated information to departments, institutions and the citizens.
- **Sensors and Connected Devices** involves capturing different kinds of signals from the environment and sending these signals to computers in management centres via networks. Sensor networks allow to obtain big volume of data and they can be used in several application areas such as smart street and traffic lights, management and quality of water resources, smart consumption, pollution and environmental risk control, emergency services and natural disaster alerts.

- **Connectivity Infrastructure** involves fixed and/or mobile data networks which includes fiber, cable and wireless (Wi-Fi, 3G, 4G, or radio) transmission networks to send and receive data.

Smart cities conducted two main components as sustainability and communication. Sustainability refers to effective use of resources and improvement of life standards. Communication refers to instant data collection and acquisition of operating systems and enabling the users to obtain more detailed information. In a sustainable smart city architecture there are basically four layers which are Sensing layer, Communication layer, Data layer and Service/Application layer (URL 2).

- **Sensing layer** involves sensors and other data capturing devices that are spread across the urban area which collect data about various thematic activities in the physical environment and send to data centres in the data layer.
- **Communication layer** enables the transmission and conversation between sensors and data processing platform in the data layer. Traditional LAN, MAN, and WAN communication technologies such as WiFi, optic fiber, ethernet, broadband powerline communication, and mobile communication technologies like UMTS and LTE are used (URL 2).
- **Data layer** can be accepted as the intelligence layer of smart-city architecture. A smart city can work effectively only with the existence of reliable and interoperable data. This layer involves data servers that process data by applying different statistical models such as predictive, descriptive and decision.
- **Service/Application layer** works as a cross-department operation centre. The citizens can access and share the information via mobile applications and web portals which are designed on this layer. Citizens can build services to improve the implementations by using this data (URL 2).

3. BIG DATA IN THE SMART CITY CONCEPT

It is aimed to explain and clarify the big data and its importance in the smart city concept. In this section, it is also examined the big data platforms and its implementations in smart city applications.

3.1 Big Data Definitions in the Smart City Concept

Along with the rapid growth and development of ICTs, the amount of storable data has increased for processors and storage areas. However, regardless of the size of the storage space, large volume of data should be processed at ever-growing speed. Nowadays, computers, smartphones, cameras, GPS, sensors, social media, games and applications, and even people themselves can be data sources (Al Nuami et al., 2015). Due to the diversity and complexity of this large volume of data, processing this data effectively and rapidly is one of the most important questions in the smart city concept (Hashem et al., 2015). Several definitions for big data concept are presented in Table 1.

Table 1. Definitions for big data

Source	***Definition***
TechAmerica, 2012	Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.
Chen vd., 2012	A complete set of data sets and analytical techniques that are large and complex enough to require advanced and unique data storage, management, analysis and visualization technologies.
Schönberger and Cukier, 2013	The phenomenon that brings together three fundamental factors for analysing, understanding and organizing information: 1. More data, 2. Incomplete (unstructured) data, 3. Correlation.
Khan vd., 2014	Highly distributed and unstructured large volume data set that exceeds the processing capabilities of traditional database engines.
SAS, 2017	Big data is a popular term used to describe the exponential growth, availability, and use of information, both structured and unstructured.
IBM, 2017	Data coming from everywhere; sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction record, and cell phone GPS signal to name a few.
Oracle, 2017	Big data describes a holistic information management strategy that includes and integrates many new types of data and data management alongside traditional data.
Gartner Inc., 2017	Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.
NIST, 2017	Big Data consists of extensive datasets –primarily in the characteristics of volume, variety, velocity, and variability– that require a scalable architecture for efficient storage, manipulation, and analysis.
Microsoft, 2017	Big data typically refers to collections of datasets that, due to size and complexity, are difficult to store, query, and manage using existing data management tools or data processing applications.

From all these definitions, the basic principles of big data are generally defined as 5Vs (White 2012; Demchenko and Membrey, 2014; Jin et al., 2015, Yin and Kaynak, 2015, Viceconti et al., 2015). These principles can be described as follows:

- **Volume:** This concept refers to the size of the data that has been produced from all the sources. Because data producing techniques are increasing day by day, technologies for archiving, processing, integrating and storing such big data should be managed and planned properly (Russom, 2011).
- **Variety:** This component refers to different types of data being produced in different environments and generally not being structural (Fosso Wamba et al., 2014). It is very important to integrate this diverse range of data and transform into one another for effective data processing and management.
- **Velocity:** This concept means that production and distribution speed of the big data is very high (Russom, 2011). Big data production speed increases every day and reaches an incredible dimension in seconds. Rapidly growing data reveals that the number and diversity of transactions needed this data also increase at the same rate.
- **Value:** Value is considered as the most important component of the big data concept and means that relevant data creates value for applications, institutions, companies etc. (Yin and Kaynak, 2015). In other words, large amount of data that are complex, diverse, and difficult to manage should be made available to the user.
- **Veracity:** Value of the big volume data depends on its accuracy. It is very important to make the necessary analysis for transition of wide range of large volume data through the right layers with right level of security and privacy (Beulke, 2011). Data must be visible or hidden by the authorized people in the right layer.

Apart from these basic principles described above, the number of principles has been increased by many authorities to seven (Khan et al., 2014, Agrahari and Rao, 2017) and to ten (Balaenand et al., 2017, Firican 2017). These additional principles in brief are:

- **Visualization:** Considering visualization is very important concept for data interpretation. Using graphs, diagrams and charts to visualize data in complex structure is more operational than using texts, tables and reports (Agrahari and Rao, 2017).
- **Validity**: Similar to the veracity, it indicates how accurate and valid the data for intended use. The critical aspect separating validity from the veracity is that data should be valid and up to date for application needs (Khan et al., 2014).
- **Volatility:** In most general terms volatility means how long the data is valid and how long it must be stored for an effective data management (Firican, 2017).
- **Vulnerability:** The big data concept has brought with some security vulnerabilities together. We all face with many security incidents, especially through social media and smartphones (Firican, 2017).
- **Variability:** Data storage volume depends on the data type and format. Due to the large data sizes resulting from a large number of data types and resources, big data sets are relatively variable (BalaAnand et al., 2017).

3.2 Big Data Solutions

Due to the characteristic features (volume, diversity, speed, etc.), it is very difficult to benefit from big data sets by using traditional processing methods and technologies (Akdamar, 2017). For this reason, various methods, algorithms and technologies have been developing to process and benefit from big data. Big data solutions can be investigated as technologies and infrastructures.

3.2.1 Big Data Technologies

One of the pioneering and most widely known big data technology is **Apache Hadoop**. The Apache Hadoop is an open source software library developed by Doug Cutting who is also the developer of the popular search engine library Apache Lucene (URL 3). It is a Java based framework which enables the distributed processing of large volume of data across several computer clusters by using basic programming models (URL 4). It is designed with the ability to scale up from single servers to thousands of machines which they offer local storage and computation ability (URL 4). Apache Hadoop provides a significant performance boost with the distributed file system and the parallel processing power (Cetin, 2014). Apache Hadoop includes the modules explained below (URL 4):

- *Hadoop Common:* The common utilities such as file system and operating system abstraction supporting the other Hadoop modules.
- *Hadoop Distributed File System (HDFS):* A highly fault tolerant distributed file system which ensures high-throughput access to data.
- *Hadoop YARN:* A framework for scheduling the jobs and managing the cluster resource.
- *Hadoop MapReduce:* YARN-based system to process large volume of data that implements the MapReduce programming model in parallel clusters.

The MapReduce programming model enables parallel process of large data sets (multi-terabyte data sets) stored in distributed architecture (Dogan and Arslantekin, 2016). The user uses map and reduce functions for defining a job and the system itself manages the required data partitioning, parallel processing and error management for this job (Demir, 2012). In the map phase, input is separated into independent small chunks that are processed in a completely parallel manner (JRC, 2014). So, it takes less time to process the split parts than to process all the input data. In this way large amounts of data can be read quickly and data processing speed and efficiency increase. In the reduce phase, intermediate product created in the map phase is transferred into the node which the reduce task is managed. The results are gathered and written to the desired source. The framework deals with planning tasks, monitoring them and re-executing the failed tasks. Since the MapReduce framework runs on a distributed architecture, map and reduce tasks are handled independently from each other.

The Hadoop ecosystem also provides an alternative execution engine to MapReduce called ***Tez*** which is a generalized data-flow programming framework (JTC, 2014). It is built on Hadoop YARN for ensuring a flexible, stable and powerful engine to operate random tasks and it manipulates data for interactive use-cases. Tez is being adopted by commercial software (e.g. ETL tools) and by other frameworks in the Hadoop ecosystem such as ***Hive*** and ***Pig***. Other infrastructures in the Hadoop ecosystem includes:

- Infrastructures for cluster and data management (URL 4): ***Ambari*** is used for preparing, managing, and tracking Apache Hadoop clusters; ***Avro*** is used for data serialization; ***Chukwa*** is used for storing and managing of large and distributed data; ***Pig*** is used for a parallel computation of big data sets by providing high order data-flow language; ***ZooKeeper*** is used for coordination services of large data sets.
- Database and data warehousing infrastructures (URL 4): ***Cassandra,*** is a scalable multi-master database; ***HBase*** as a scalable, distributed database; ***Hive*** as a data warehouse infrastructure; ***Drill*** as a query engine that does not use MapReduce and support low latency queries natively on rapidly evolving multi-structure datasets at scale.
- Database and data warehousing infrastructures (URL 4): ***Mahout*** as a scalable machine learning and data mining library; ***Spark*** as a fast and general compute engine for Hadoop data.

3.2.2 Big Data Infrastructures

Big data infrastructures can be examined as databases which are responsible for data storage and processing at a low level and analytical tools that are used for data processing for higher level tasks. There are two broad categories of databases used in big data concept: NoSQL and NewSQL.

NoSQL Databases: NoSQL (not only Structured Query Language-SQL) systems are emerged as an alternative to Relational Database Management System. NoSQL databases ensure the storing and retrieving of data and they are designed differently than the tabular relations in traditional databases (JRC, 2014) and allows searches from different structure and sizes of data (Dogan and Arslantekin, 2016). A common advantage of the NoSQL databases when compared to the traditional relational databases is that the NoSQL databases scale horizontally (or scale out), while the traditional databases scale up only (Pokorny, 2013). NoSQL databases use *data sharding* technique for scaling out by horizontal data partitioning (JRC, 2014).

On the other hand, better performance and scalability impose a small sacrifice in terms of consistency. The relational databases have the **ACID** properties which shortly means **A**tomicity, “all or nothing”; **C**onsistency; “the result of each transaction are tables with legal data”, **I**solation, “transactions are independent” and **D**urability, “database survives system failures”, respectively (Pokorny, 2013). But in a NoSQL database, it is hard to achieve the ACID properties synchronously. In the NoSQL databases, the sense of consistency refers the eventual consistency (JRC, 2014). Accordingly, the **BASE** set of properties (**B**asically

Available, **S**oft state, **E**ventual consistency) has been specified for the NoSQL databases (Pritchett, 2008).

NoSQL is interpreted as **N**ot **o**nly **S**QL, for emphasizing that a NoSQL database may also support an SQL-like query language. There are different types of NoSQL databases. The most common NoSQL database types are *key-value stores, column databases, document stores, graph databases* and *array databases* (JRC,2014):

- *Key-Value Stores:* The Key-value (KV) stores are the simplest NoSQL databases using the associative array. Data is represented as key-value pair collections. The database contents are organized in rows and every item in the KV databases is kept as an attribute name (a key), along with a value. Examples of KV stores are the *redis*, *Voldemort*, *riak*, etc.
- *Column Databases:* Data is stored as combination of key/value pairs grouped into collections. New attributes (columns) can be inserted and the key-value pairs can be stored in a massively parallel system (JRC, 2014). The column databases support both horizontal and vertical partitioning and high availability. Examples of column databases are the Hadoop databases *HBase* and *Cassandra*, Google databases *BigTable* and *BigQuery*, *Impala* etc.
- *Document Stores:* In the document store concept, documents encapsulate and encrypt data in specified standard formats such as *JavaScript Object Notation (JSON)* and the *eXtensible Markup Language (XML)* or encodings (JRC, 2014). The documents of a document store are identified by unique keys, an API or query language to retrieve documents based on their contents (JRC, 2014). Popular document stores include *MongoDB, CouchBase, Apache CouchDB, eXist* etc.
- *Graph Databases:* Graph databases use Create, Read, Update, and Delete (CRUD) methods by indicating graph data model (Robinson et al., 2013). It is an online database management system and its main focus on graph database use is on social network analysis. Examples of popular graph databases are the *Neo4j, InfoGrid, Oracle Spatial and Graph, Oracle NoSQL Database* etc.
- *Array Databases:* The array databases provide database services specifically for arrays representing sensor, simulation, image, or statistics data. The main purpose of an array database is to access and manipulate to large arrays and sub-arrays. Most popular examples are, the *Oracle Spatial and Graph, MonetDB, PostGIS, rasdaman, SciDB* etc (JRC, 2014).

NewSQL Databases: NewSQL can be accepted as modern relational database systems for providing the similar scalable performance of NoSQL for online transaction processing (OLTP) read-write works and at the same time maintaining a relational database system's ACID guarantees JRC, 2014). The NewSQL databases can be classified as:

- *New architectures:* These new database platforms are modelled to manipulate in distributed clusters of shared-nothing nodes, where each node has a subset of data (JRC, 2014). Examples of NewSQL databases are *Google Spanner, FoundationDB, ActorDB, Clustrix, MemSQL, SAP HANA, VoltDB, Trafodion,* etc.

- *SQL Engines:* The systems of this category are equipped with highly optimized storage engines for SQL, which provide the SQL programming interface (JRC, 2014). Examples include *TokuDB* and *InfiniDB.*
- *Transparent sharding:* These systems provide a *database shard* which is a horizontal partition of data in a database or search engine, to split databases across multiple nodes automatically. Examples of the systems are *dbShards, Scalearc*, and *ScaleBase*. (JRC, 2014).

3.3 Example Applications in the Smart City Concept

Big data technologies promote effective and successful data management, storage and processing capabilities to generate information improving various smart city services. These examples can be accepted as successful smart city applications. Table 2 reviews that how big data applications are utilized in smart city projects around the world.

Table 2. Successful Smart City Projects

Smart city component	***Big Data Project***	***Location***
Traffic and Environmental Issues	Stockholm implemented smart city applications to solve the problems caused mainly by environmental and traffic issues. More than half million entries about waste fractions, locations and weights were collected and actualized. These large amounts of data about waste management was used for collecting and examining to define the inadequacies in waste collection routes which cause traffic within the city (Shahrokni et al., 2014).	Sweden
	In Singapore, OneMap portal which can be openly accessed on the internet contains free up-to-date information POIs such as WiFi hotspots, public services, road conditions, national monuments etc. By using OneMap citizens can compute distances between potential targets or access instant traffic information (Tao, 2013).	Singapore
Healthcare	Ministry of Health and Welfare in South Korea initiated the Social Welfare Integrated Management Network with the purpose of managing welfare benefits and services to citizens provided by the central and local government. Within the project 385 different types of public data from 35 agencies were analysed (Kim et al., 2014).	South Korea
Safety	Singapore government initiated the Risk Assessment and Horizon Scanning (RAHS) program to identify national security threats, infectious diseases, and other national concerns in 2004. Within the project large volume data sets were examined for mainly managing interior security threats like infectious diseases, terrorist attacks and economic distresses (Kim et al., 2014).	Singapore
Natural Resources & Energy	In 2004 UK government instituted the Horizon Scanning Centre (HSC) in order to enhance the ability of the government for dealing with multi-disciplinary challenges. In addition, the HSC's Foresight International Dimensions of Climate Change effort examined climate change and its effects about the availability of international stability and security, food and water	UK

Smart city component	*Big Data Project*	*Location*
	reserves and regional tensions by operating advances analysis on multiple data channels in 2011 (Kim et al., 2014).	
Public Administration	With the Helsinki Region Infoshare more than 1030 databases which cover a wide range of urban phenomena such as transport, economics, conditions and employment were made available to individuals, government, academy, business and research institutions. The project promotes the opportunity of public involvement in decision-making process.	Finland
Government & Agency Administration	USA initiated data.gov project to ensure as a part of the transparent and accountable government policy in 2009. In the project a warehouse which contains 420,894 datasets about health care, economy, transportation, education, and human services was designed (Neirrotti et al., 2014). In Syracuse, NY, government initiated a project called "Smarter City" in 2011 together with IBM, in order to use large volume data to determine and prevent vacant residential properties. Michigan's Department of Information Technology established a data warehouse for ensuring a single source where information was collected and managed (Neirrotti et al., 2014).	USA

4. CONCLUSION

Data sets coming from various data sources are managed in relational database environment traditionally. In Urban GIS applications, cadastral maps, zoning plans and other planning maps, taxation data, and infrastructure data can be related, updated, and queried in traditional multiuser GIS databases. Data producing techniques are developing as time goes by and big data needs to be used in smart city applications. For example, traffic and transportation applications should use real time traffic and velocity data from sensors, beside road network dataset. Reviewing smart city applications show how big data analysis is important for successful urban management. Analysing and recognizing the patterns especially in traffic, healthcare and safety issues, providing detailed and immediate data solutions or integration and collaboration of government agencies and citizens in decision making helps to reach successful smart city. However, some characteristics especially in technological manner are still not well defined and assimilated in smart city concept. Right tools and methods, therefore, should be used for big data processing and for achieving successful applications and advance services in smart cities.

In smart city applications, different data types and technologies are required according to the target application domain. Therefore, all aspects should be analysed and feasible investment strategies should be determined by decision makers. It is required to test the usability of NoSQL structured databases for document, array, and column-based data types. Besides using relational database structure on Urban GIS applications, text, images, and real time sensor data should be processed in the applications suitable to big data principles. Hadoop ecosystem with its components or similar platforms should be tested for urban analytics applications that need geographic information or parallel processing capabilities.

ACKNOWLEDGMENTS

This work was supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under project no. 116Y204

REFERENCES

Agrahari, A., Rao, D., 2017. Big Data: Technologies, Tools and Trends. nternational Research Journal of Engineering and Technology (IRJET), 4 (10), 640-649.

Akdamar, E ., 2017. Akıllı Kent İdealine Ulaşmada Açık Verinin Rolü. Sosyal Bilimler Araştırma Dergisi, 6 (1), 45-52.

Alawadhi S. et al. (2012) Building Understanding of Smart City Initiatives. In: Scholl H.J., Janssen M., Wimmer M.A., Moe C.E., Flak L.S. (eds) Electronic Government. EGOV 2012. Lecture Notes in Computer Science, vol 7443. Springer, Berlin, Heidelberg.

Al Nuaimi, E., Al Neyadi, H., Mohamed, N., ve Al-Jaroodi, J., 2015. Applications of big data to smart cities. Journal of Internet Services and Applications, 6(1), 1–15.

BalaAnand, M., Karthikeyan, N., Karthik, S., Sivaparthipan, C. B., 2017. A Survey on BigData with Various V's on comparison of Apache Hadoop and Apache Spark. Advances in Natural and Applied Sciences, 11(4), 362-369.

Beulke, D., 2011. Big Data Impacts Data Management: The 5 Vs of Big Data

Cetin, Y., 2014. MapReduce Kullanarak Rdfs Üzerinde Dağıtık Çıkarsama. McS Thesis, TOBB Economy and Technology University.

Chen, H., Chiang R., and Storey V., 2012. Business intelligence and analytics: From big data to big impact. MIS Quarterly: Management Information Systems, 36, 1165-1188.

Chen, M., Mao, S., Zhang, Y., Leung, V. C. M., 2014. Big Data. SpringerBriefs in Computer Science. Cham: Springer International Publishing.

Chourabi, H., Nam, T., Walker, S., Gil-Garcia, J. R., Mellouli, S., Nahon, K., 2012. Understanding Smart City initiatives: An integrative framework. In 45th Hawaii international conference on system sciences, 4–7 Ocak, Maui, HI.

Correia, L. M., ve Wünstel, K., 2011. Smart Cities applications and requirements. White Paper of the Experts Working Group, NetWorks European Technology Platform.

Demchenko, Y., Membrey, P., 2014. Defining Architecture Components of the Big Data

Ecosystem.

Demir, I., 2012. Hadoop Tabanlı Büyük Ölçekli Görüntü İşleme Altyapısı. McS Thesis, Kocaeli University.

Dogan, K., Arslantekin, S., 2016. Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum. TCF, 56 (1), 15-36.

Firican, G., 2017. The 10 Vs of Big Data. https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D., 2014. How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. Int. J. Production Economics, 165, 234-246.

Gartner INC., 2017. http://www.gartner.com/it-glossary/bigdata/

George, G., Haas, M. R., Pentland, A., 2014. Big Data and Management. Academy of Management Journal, 57(2), 321–326.

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Ullah Khan, S., 2015. The rise of "big data" on cloud computing: Review and open research issues. Information Systems, 47, 98–115.

Hollands, R. G., 2008. Will the real smart city please stand up? Intelligent, progressive or entrepreneurial? City, 12 (3), 303-319.

Inter-American Development Bank (IDB), 2016. The Road toward Smart Cities: Migrating from Traditional City Management to the Smart City.

Internet of Things Global Standards Initiative, 2012. http://www.itu.int/en/ITU T/gsi/iot/Pages/default.aspx.

Jin, X., Wah, B. W., Cheng X., Wang, Y., 2015. Significance and Challenges of Big Data Research. Big Data Research, 2, 59–64.

JRC (Europian Commision-Joint Research Centre), 2014. Scientific Pilot Study on Big Data Analaysis.

Khan, M. A., Uddin, M. F., Gupta, N., 2014. Seven V's of Big Data: Understanding Big Data to extract Value. Proceedings of 2014 Zone 1 Conference of the American Society for Engineering Education.

Kim, G. H., Trimi, S., Chung, J. H., 2014. Communications of the ACM, 57 (3), pp. 78-85.

McAfee, A., & Brynjolfsson, E., 2012. Big Data's Management revolution. Harward Business Review.

McKinsey Global Institute, 2011. Big Data: The next frontier for innovation, competition and productivity.

Microsoft, 2017. Developing big data solutions on Microsoft Azure HDInsight.

Nam, T. A., Pardo, T., 2011. Conceptualizing Smart City with Dimensions of Technology, People, and Institutions, 284-285.

Neirotti, P., De Marco, A., Cagliano, A., Mangano, G., Scorrano, F., 2014. Current trends in Smart City initiatives: Some stylised facts. Cities, 38, 25-36.

NIST, 2017. Big Data Public Working Group, Big Data Interoperability Framework: Definitions.

Official Webpage of IBM Company: http://www.ibm.com/software/data/bigdata/

Official Webpage of Oracle: https://www.oracle.com/big-data/guide/what-is-big-data.html.

Official Webpage of SAS Company: http://www.sas.com/big-data/

Pokorny, J. 2013. "NoSQL databases: a step to database scalability in web environment", IJWIS 9(1): 69-82.

Pritchett, D., 2008. BASE: An Acid Alternative. Queue - Object-Relational Mapping, 6 (3), pp. 48-55.

Robinson, I., Webber, J., Eifrem, E., 2013. Graph Databases. O'Reilly Media.

Russom, P., 2011. The Three Vs of Big Data Analytics: TDWI.

Sagiroglu, S., Sinanc, D., 2013. Big data: A review. Collaboration Technologies and Systems (CTS). International Conference, 42–47.

Schönberger, V. M., Cukier K., 2013. Büyük Veri - Yaşama, Çalışma ve Düşünme Şeklimizi Dönüştürecek Bir Devrim. Istanbul: Paloma.

Schaffers, H., Komninos, N., & Pallot, M., 2012. Smart Cities as Innovation Ecosystems, Sustained by the Future Internet. FIREBALL White Paper.

Shahrokni, H., Van der Heijde, B., Lazarevic, D., & Brandt, N., 2014. Big data GIS analytics towards efficient waste management in Stockholm. Paper presented at the Proceedings of the 2014 Conference ICT for Sustainability.

Tao, W., 2013. Interdisciplinary urban GIS fors mart cities: advancements and opportunities. Geo-spatial Information Science, 16 (1), 25-34.

TechAmerica Foundation's Federal Big Data Commission, 2012. Demystifying big data: A practical guide to transforming the business of Government. http://www.techamerica.org/Docs/fileManager.cfm?f=techamericabigdatareport-final.pdf.

Terzi, F., Ocakcı, M., 2017. Kentlerin Geleceği: Akıllı Kentler, İTÜ Vakfı Dergisi, 77, 10-13.

Urban Sustainability Directors Network (USDN), 2015. Getting Smart About Smart Cities. USDN Innovation Fund

Viceconti, M., Hunter, P., Hose, R., 2015. Big Data, Big Knowledge: Big Data for Personalized Healthcare. IEEE Journal of Biomedİcal And Health Informatics. 19 (4), 1209-1215.

Washburn, D., Sindhu, U., Balaouras, S., Dines, R. A., Hayes, N. M., & Nelson, L. E., 2010. Helping CIOs understand ''Smart City'' initiatives: Defining the Smart City, its drivers, and the role of the CIO. Cambridge, MA: Forrester Research, Inc.

White, M., 2012. Digital workplaces: vision and reality. Bus. Inf. Rev. 29 (4), 205–214.

Yin, S., Kaynak O., 2015. Big Data for Modern Industry: Challenges and Trends. Point of View, 103, (2), 143-146.

Web References

URL 1. https://esa.un.org/unpd/wpp/

URL 2. https://itunews.itu.int/en/5217-smart-sustainable-cities-a-road-map.note.aspx

URL 3. https://www.b3lab.org/hadoop-nedir/

URL 4. http://hadoop.apache.org/

BIOGRAPHICAL NOTES

Arif Cagdas AYDINOGLU

Gebze Technical University
Geomatics Engineering Department
41400 Gebze/Kocaeli
Turkey
Tel. +90 216 650 1803
Email: aydinoglu@gtu.edu.tr

Rabia BOVKIR

Gebze Technical University
Geomatics Engineering Department
41400 Gebze/Kocaeli
Turkey
Tel. +90 216 650 1813
Email: rbovkir@gtu.edu.tr