ISSN No. 2456-0472
The Engineering Journal of Application & Scopes, Volume 1,Issue 2, July 2016 Page: 94-100

THE BIG DATA APPROACH TOWARDS SMART CITIES

Supriya shukla
Department of Computer Science Engineering, College of Engineering Roorkee, Roorkee, INDIA
Email: supriyaluthra@yahoo.co.in

Abstract
Today, Indian government is adopting the smart city concept in many of the cities and implementing Big Data applications that support smart city components to reach the required level of sustainability and improve the living standards. Big Data has exaggerated attention from academician, industrialist as well as government. It is a very crucial research area. Big Data is term defining collection of large and complex data sets that are difficult to process using conventional data processing tools. Every day, we create trillions of data all over the world. These data is coming from social networking sites, scientific experiments, mobile conversations, sensor networks and various other sources. We require new tools and techniques to organize, store, process and analyze Big Data. This paper systematically presents the importance of Big Data applications in smart cities and explores the applications, benefits, challenges and solutions of incorporating Big Data applications for smart cities.

Keywords: Big Data, smart cities, Hadoop, HDFS.

I. Introduction

In digital world, IBM estimates that every day 2.5 quintillion bytes of data are created, out of which 90% of the data in today has been created in the last two years .for example data gather from sensors used to collect information regarding climate, social media sites posts, digital pictures and audios/videos uploaded on internet, purchase transaction records, and cell phone conversations, etc.

In general, it refers to the collection of large and complex datasets which are difficult to process using traditional database management tools or data processing applications. These are available in structured, semi-structured, and unstructured format in petabytes and beyond.

Another useful perspective is to characterize Big Data as having high volume, high velocity, high Variability and high variety [1]:

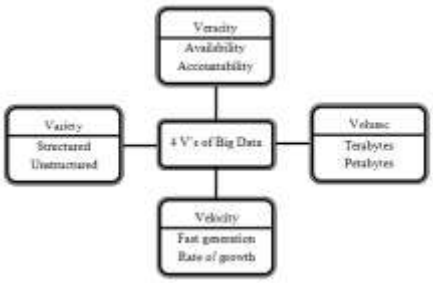

Figure 1: Chracteristics of Big Data

(i)Volume – The name 'Big Data' itself contains a term which is related to size .it is the quantity of data that is generated, it is the size of the data which determines the value and potential of the data under consideration and whether it can actually be considered as Big Data or not.

(ii) Variety- This means that the category to which Big Data belongs to is also a very essential fact that needs to be known by the data analysts so that they can effectively use the data.

(iii) Velocity- The term 'velocity' in this context refers to the speed of generation of data or how fast the data is generated and processed to meet the demands and this challenge lies in the path of growth and development of Big Data.

(iv) Variability- This refers to the inconsistency which can be shown by the data at times, thus restrict the process of being able to handle and manage the data effectively.

II. Smart City

ISSN No. 2456-0472
The Engineering Journal of Application & Scopes, Volume 1,Issue 2, July 2016 Page: 94-100

Smart Cities Mission is an urban revitalization and furnishing program by the Government of India with a mission to develop 100 cities all over the country making them citizen friendly and sustainable.[3] The Union Ministry of Urban Development is responsible for implementing this mission in collaboration with the state governments of the respective cities. The government of India under Prime Minister has a vision of developing 100 smart cities as satellite towns of larger cities and by modernizing the existing mid-sized cities.

The 100 potential smart cities were nominated by all the states and union territories based on Stage 1 criteria, prepared smart city plans which were evaluated in stage 2 of the competition for prioritizing cities for financing. In the first round of this stage, 20 top scorers were chosen for financing during 2015-16. The remaining will be asked to make up the deficiencies identified by the Apex Committee in the Ministry of Urban Development for participation in the next two rounds of competition. 40 cities each will be selected for financing during the next rounds of competition.

Figure 2: 20 smart cities of India

III. Benefits and opportunities

Many opportunities are made possible by using Big Data analytics in smart city applications. Therefore, in this section reflect the benefits and opportunities that will help in making the decision to convert or redesign a city to become a smart city[2]. Some of the benefits of having a smart city include the following:

A. Efficient resource utilization: Now a day, many resources are becoming either scarce or very expensive; it is important to integrate better and controlled utilization of these resources by the use of the technological systems such as Enterprise resource planning (ERP) and Geographic Information System (GIS) [3]. With monitoring systems at work, it will be easier to spot waste points and better distribute resources while controlling costs, and reducing energy and natural resources consumption. The important aspects of smart city applications is that they are designed for interconnectivity and data collections which can also facilitate better collaboration across applications and services.

B. Better quality of life: With better services, more efficient work and living models, and less waste (in time and resources), smart city citizens will have a better quality of life. This is the result of better planning of living/work spaces and locations, more efficient transportation systems, better and faster services, and the availability of enough information to make informed decision.

C. Higher levels of transparency and openness: The need for better management and control of the different smart city aspects and applications will drive the interoperability and openness to higher levels, this will increase information transparency. This will encourage collaboration and communication between entities and creating more services and applications that further enhance the smart city.

The opportunities to achieve these benefits are available; however, they require investing in more technology, better development efforts and effective use of Big Data. There is also the need to set policies to ensure data accuracy, high quality, high security, privacy, and control of the data as well as using data documentation standards to provide guidance on the content and use of the datasets [10]. These benefits to be achieved require high levels of sophistication

and involvement in terms of the applications, resources and people involved

This also helps in real-time solutions to challenges in agriculture, transportation, and crowd management as applications and systems are integrated and information flows easily cross applications and entities [10]. There are many examples of Big Data applications serving smart cities such as:

A. **Smart education** [4]: Big Data in education is generated mainly by collecting data on people (e.g. students, teachers, parents, administrators, and other support personnel), infrastructures (e.g. schools, libraries, computing facilities, educational locations, museums, universities, and other related entities), and information (e.g. courses, books, exams, grades, economic surveys, assessments, reports, and much more). This data can create a useful resource for analysis and extracting useful trends, models and using them to offer better and more enhanced education.

Big Data provides a solution to enhance the education efficiency, effectiveness, and productivity using education smart services that are flexible and intelligent to provide better use of information, enhanced control and assessment, higher support for life-long learning for all people (citizens and stakeholders). Smart education applications will engage people in active learning environments that allow them to adapt to the rapid changes of society and the environment. In addition, by relying on Big Data collected in the field and correctly processed to generate the required information, we will have a positive effect on the knowledge levels and teaching/learning tools to deliver or acquire knowledge.

B. Smart traffic lights [5]: One of the main aspects of smart cities is a good control of the traffic flow within the city, which will enhance the transportation systems and improve the citizens' commutes and the cities overall traffic patterns. When the population increases, traffic problems, pollution, and economic problems happen. Due to this, the use of smart traffic lights and signals is one of the most important techniques that smart cities use to deal with high volumes of traffic and congestions. Smart traffic lights and signals should be interconnected across the traffic grids to offer more information about traffic patterns. Each sensor detects a different parameter of the traffic flow (e.g. the speeds of cars, traffic density, waiting time at the lights, traffic jams, etc.). The system makes decisions according to the values of these parameters and gives the appropriate instructions to the lights and signals. Thus, the more data available to this system, the more informed decisions it will be able to make. As a result, to offer the best possible services in smart traffic lights.

C. Smart grid: The smart grid is an important component of a smart city. It is a renovated electrical grid system that uses information and communication technology to collect and act on available data, such as information about the behaviors of suppliers and consumers, in an automated fashion to add some values [17]. It improves the efficiency, reliability, economics, and sustainability of the production and distribution of electric power. A smart grid uses computer-based remote controls with two-way communication technology between power producers and consumers to increase grid efficiency and reliability through system self-monitoring and feedback. This involves placing smart sensors and meters on production, transmission, and distribution systems in addition to consumers' access points to get granular near real time data about the current power production.

IV Challenges

There are many challenges in designing, developing and deploying of Big Data applications for smart cities. There are also some controversies related to the definition, use and benefits of Big Data for smart cities. These relate to available Big Data tools, real-time analytics, accuracy, representation, cost, and accessibility. Such issues can affect the performance of smart city applications and services relying on Big Data [6], here we will address some of the key challenges in using Big Data in smart cities.

A. Data sources and characteristics: Data is generated from many different sources in many different formats. There are a lot of new data formats many of which are unstructured (e.g. images, audio, tweets, video, server logs, etc.). This data need to be managed and classified into a structured format using some form of advanced database systems [7].

Many identified different Vs of Big Data the most agreed upon are the 3 Vs: Velocity, Volume and Variety. Several more were added such as Validity, Veracity, Volatility and Value [20] as well as Variability [2]. Just trying to encompass these different attributes of Big Data generates very complex models and approaches and makes it hard to manage. This is simply because the current methodologies or data mining software tools cannot handle the large size and complexity. In addition, there are some challenges that may be faced in the future, such as analytics architecture, evaluation, distributed mining, time evolving data, compression, visualization, and hidden Big Data [8]. When considering smart city applications utilizing Big Data difficulties arise in various areas. For one, collecting the data by itself is complicated by the existence of multiple sources with different formats and types and different usage and access policies. In addition, the unstructured nature of the data makes it hard to categorize and organize and an easily accessible way for applications to use.

B. Data and information Sharing: Sharing data and information among different city departments is another challenge. Each government and city agency or department typically has its own warehouse of confidential or public information. Most of which are often reluctant to share what might be considered proprietary data. In addition, some data may be governed by certain privacy conditions that make them hard to share across different entities.
The challenge here is to make sure not to cross the fine line between collecting and using Big Data and ensuring citizens' rights of privacy [8]. This is applicable within any smart city since there are many sectors and industries involved. Smart city applications will need to find ways to prevent or reduce the barriers to achieve seamless information sharing and exchange among different entities [8].

C. Data Quality: Looking at more fundamental aspects of Big Data, there are a number of challenges that are associated with the quality of the data. Data captured by different people under special regimes and stored in distinctive databases is rarely stored in any standard formats [9]. Relying on crowd sourcing and collaboration of multiple providers will result in data that suffers from a lack of structure and consequently consistency, heterogeneity, and disparity issues will have a greater chance to occur.

D. Security and privacy: Another one of the major challenges in a smart city and with using Big Data is the security and privacy issues. In basic terms this mean that databases may include confidential information related to the government and people, so they need high levels of security policies and mechanisms to protect this data against unauthorized use and malicious attacks. In addition, smart applications integrated together across agencies also require high security since the data will move over various types of networks, some of which may be pen or unsecure [8].

Although specific smart city entities can claim ownership of most Big Data, a lot of it includes personal and private information about individuals. Health and medical records, financial and bank records, retail history, and much more all provide intimate views of the people they represent. Many view access to this type of data as a violation of a person's legal rights for privacy. Making sure that stringent privacy policies are put in place and properly enforced represents a major challenge for Big Data smart city applications developers and users.

E. Cost: Cost is a sensitive subject that involves the ways public authorities may affect people when they use these solutions. For example, using an energy usage reduction system [10], which forces the government to use new systems, components or features to monitor consumption and record information. This leads to creating a smart energy management system; however, it is also a very expensive to implement [10]. In addition if such a project is not implemented correctly from the beginning, it might cause a big problem, result in very high costs, and the city may be negatively affected.

V Requirements

This section will cover the key components required to design and implement smart city applications utilizing Big Data components. Data collection and capturing from sensors, users, electronic data readers and many others pose the first issue to handle as the volume rapidly grows. Storing, organizing and processing this data to generate useful results in the next issue. Fundamentally, to have effective solutions, it is required to select a number of design and development priorities in a planned manner, for example flexible design, quick deployment, achieving more thorough sense, more comprehensive interconnections, and more intelligence [10].
To further complicate the issues, handling interconnected communication infrastructures to access contextual information in smart city applications and physical spaces to support good decision making processes requires attention to various aspects of connectivity, security and privacy [2].
In this section we attempt to discuss several of these requirements to provide a general guideline for the design and development efforts. These requirements are identified based on the type of Big Data applications and the challenges of implementing these applications for smart cities. Some of these requirements are technological while others are related to citizens' awareness and governments' roles. Furthermore, some of these requirements are general and apply to any Big Data application, while others are specific to the special needs of smart city environments.

A. Big Data Management (BDM): This is where data management disciplines, tools, and platforms (both old and new) are applied to the management of Big Data (in the base definition or the extended one). Traditional data and new Big Data can be quite different in terms of content, structure, and intended use, and each category has many variations within it. To accommodate this diversity, software solutions for BDM tend to include multiple types of data management tools and platforms, as well as diverse user skills and practices.

It also includes development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs throughout its use in smart city applications. As the data comes from different sources with different formats, there is a need for advanced data management features that will lead to recognizing the different formats and sources of data, structuring, managing, classifying, and controlling all these types and structures. Big Data management for smart city applications should also provide scalable handling for massive data to support offline applications as well as low latency processing to serve effectively in real-time applications.

B. Big Data Processing Platforms: There are several Big Data platforms available with different characteristics and choosing the right platform requires an in-depth knowledge about the capabilities of all these platforms [1]. Especially, the ability of the platform to adapt to

increased data processing demands plays a critical role in deciding if it is appropriate to build the analytics based solutions on a particular platform.

Big Data applications for smart cities need to perform data analytics that usually require huge processing capability. This leads to the need for scalable and reliable software and hardware platforms. The software platforms for smart cities should offer high performance computing capabilities, be optimized for the hardware being used, is stable and reliable for the different data-intensive applications being executed, supports stream processing, provides a high-levels of fault resilience, and is supported by a well-trained and capable team and vendor. There are different available software platforms for Big Data analytics such as Hadoop,Spark, Cloudera Manager,Hbase,etc.

These platforms work well on cluster systems that can provide a powerful and scalable hardware platform to meet the requirements of Big Data applications for smart cities.

C . Smart network infrastructure: Most Big Data applications for smart cities require to have smart networks connecting their components including residents' equipment such as cars, smart house devices, and smart phones. This network should be capable of efficiently transferring collected data from their sources to where Big Data is collected, stored, and processed and to transfer responses back to the different entities that need them in the smart city.

D. Advanced Algorithms: Standard algorithms used in regular applications may not be sufficient or efficient enough to handle Big Data applications due to their unique requirements and pressing need for high volume high speed processing. For example, most available data mining algorithms are not very suitable for Big Data mining applications as their design is based on limited and well defined data sets [11]. Big Data applications for smart cities will need to implement advanced and more sophisticated algorithms to deal with Big Data efficiently. Some of these algorithms need to be designed for real-time application support while others can be designed for batch or offline processing. These algorithms need to be optimized to handle high data volumes, large variety of data types, time constraints on decision making processes, and distributed components across various geographical locations.

E. Security and Privacy:

Security and privacy concerns are growing as Big Data becomes more and more accessible. The collection and aggregation of massive quantities of heterogeneous data are now possible. Large-scale data sharing is becoming routine among scientists, clinicians, businesses, governmental agencies, and citizens. However, the tools and technologies that are being developed to manage these massive data sets are often not designed to incorporate adequate security or privacy measures, in part because we lack sufficient training and a fundamental understanding of how to provide large-scale data security and privacy. We also lack adequate policies to ensure compliance with current approaches to security and privacy.[13]

Big Data applications designers and developers must include security and privacy policies and procedures as an integral part of the design and implementation of their applications. Clear guidelines and requirements must be identified from the various users to be enforced in the applications.

F. Citizen Awareness: Citizens must be aware of how to use these solutions for smart city correctly and safely. Their active participation in providing information related to the different issues they may encounter with smart city applications will help in enhancing the quality of collected data and the performance of the applications. As a result, more effective decisions can be made from collected Big Data to enhance different smart city components [13].

Another important aspect in citizen awareness is their knowledge and practice of good safety, security and privacy practices. Adequate training and awareness campaigns need to be done to make sure that people are aware and capable of protecting their own data and environment.

VI Conclusion

It is every Indians dream that India should stand next/ ahead of developed countries. We are proving ourselves in technology intensive industries like space, defense etc. We should further leverage our abilities to build smart cities to handle the ever growing migration to cities and depleting natural / manmade resources. The only answer to address this problem is implementing smart cities.

This paper talks about smart city and Big Data ,these two modern and important concepts need to be integrated to develop smart city applications that will help reach sustainability, better resilience, effective governance, enhanced quality of life, and intelligent management of smart city resources.

Here, we quickly moves to key opportunities available with Big Data can be successfully deployed which will result in building smart applications capable of utilizing all available data to enhance their operations and outcomes. We also discussed the various challenges in this domain, along with remediation measures.

References

1. Eiman Al Nuaimi1, Hind Al Neyadi1, Nader Mohamed2* and Jameela Al-Jaroodi3
   Pantelis K, Aija L. Understanding the value of (big) data. In Big Data, 2013 IEEE International Conference on IEEE; 2013. pp. 38–42.
2. Khan Z, Anjum A, Kiani SL. Cloud Based Big Data Analytics for Smart Future Cities. In Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. IEEE Computer Society; 2013. pp. 381–386.
3. Bertot JC, Choi H. Big Data and e-government: issues, policies, and recommendations. In proceedings of the 14th Annual International Conference on Digital Government Research. ACM; 2013. pp. 1–10.
4. Tantatsanawong P, Kawtrakul A, Lertwipatrakul W. Enabling future education with smart services. In SRII Global Conference (SRII), 2011 Annual IEEE; 2011. pp. 550–556.
5. Aguilera G, Galan JL, Campos JC, Rodríguez P. An Accelerated-Time -Simulation for Traffic Flow in a Smart City. FEMTEC. 2013; 2013:26.
6. Fan W, Bifet A. Mining Big Data: current status, and forecast to the future. ACM SIGKDD Explor Newsl. 2013; 14(2):1–5.
7. Michalik P, Stofa J, Zolotova I. Concept definition for Big Data architecture in the education system. In Applied Machine Intelligence and Informatics (SAMI), 2014 IEEE 12th International Symposium on 2014. pp. 331–334.
8. Su K, Li J, Fu H. Smart city and the applications. In Electronics, Communications and Control (ICECC), 2011 International Conference on IEEE; 2011. pp. 1028–1031.
9. Lee CH, Birch D, Wu C, Silva D, Tsinalis O, Li Y, Guo Y. Building a generic platform for big sensor data application. In Big Data, 2013 IEEE International Conference on IEEE; 2013. pp. 94–102.
10. Kramers A, Höjer M, Lövehagen N, Wangel J. Smart sustainable cities–Exploring ICT solutions for reduced energy use in cities. Environ Model Software. 2014;56:52–62.
11. Mohamed N, Al-Jaroodi J, "Real-time Big Data analytics: Applications and challenges," High Performance Computing & Simulation (HPCS), 2014 International Conference on, vol., no., 2014. pp. 305,310.
12. Eiman Al Nuaimi, Hind Al Neyadi, Nader Mohamed and Jameela Al-Jaroodi.Applications of Big Data to smart cities, Al Nuaimi et al. Journal of Internet Services and Applications (2015) 6:25.