

Heterogeneous Big Smart data in Perspective of Smart Cities Development

Monika Saxena

Assistant Professor, Department of Computer Science,
Banasthali Vidyapith, Rajasthan
muskan.saxena@gmail.com

Dr. C.K. Jha

Associate Professor, Department of Computer Science,
Banasthali Vidyapith, Rajasthan
ckjha1@gmail.com

Abstract— We are living in the age of brilliance, where all things even smallest are going to be smart. In this era, smart cities are developing to enhance the quality of living of all people. That will reduce the consumption and cost of resources. In a smart city, transport energy, water industry, health care departments, agriculture departments and waste management are expected to perform intelligently, smartly and automatically. However, the size of data is always increasing and meaningless data storage is just a waste of the storage space and time. Large amount of heterogeneous data have challenges like complexities, securities and risks to the privacy of data. We need to fetch ‘Smart data’ from the big complex Data. Big data analytics and computing plays an important role in smart cities development and management. The data generated by smart cities will analysis using Big Data analytics. In this paper, we are discussing about different aspects of smart cities and their issues and challenges. Further the role of Big Data analytics and their tools and technologies.

Keywords- Smart Cities, Big Data analytics, heterogeneous data, Internet of Things

I. INTRODUCTION

The Big Data is an aspect of a smart city and on people’s lives [1]. Big data is growing; As a result, many IT companies and governments have started programs to utilize Big Data to support the sustainability and development of smart cities. That gives solutions to cities to maintain the requirements of smart city characteristics. These characteristics include transparent governance, improved quality of life, smart management of natural resources, smart traffic management, smart education, smart health facilities, development of urban area and many more. [2]. as we know, to facilitate all such services large storage and computational facilities are needed. Big Data represented large volume data with diverse and heterogeneous dimensionalities. We need to extract useful knowledge and behavior that must process called ‘**Smart data**’ from the big complex Data. Smart data is meaningful and filtered data from Big Data. In this paper we are going to discuss about requirements of smart cities their issues and challenges in terms of Big Data.

Big data analytics and computing plays an important role in smart cities development and management. The data generated by smart cities will analysis using Big Data analytics. Big Data concerns “large-volume, complex and growing data sets”. One of the basic characteristics of the Big Data represented large volume data with diverse and heterogeneous dimensionalities. For small scale data mining tasks, a single computer with hard disk and processors are enough. And for Big Data mining, because data capacity is far beyond from a single personal computer can handle, a typical Big Data processing framework will rely on cluster computers with a high computing platform, where a data mining task is deployed by running some parallel programming tools, such as SAMOA (Scalable Advanced Massive Online Analysis) or MapReduce, on a large number of computing nodes (i.e. clusters). Distribution of Big Data gives better performance in terms of variety and large-volume, distributed pattern mining is one of the best method for Big Data performance[20][22]

II. REQUIREMENTS

A. Smart Cities:

Different type of devices needed in Internet of Things:

- 1) Sensors used to monitor utilities such as water supply, Electricity, transport to get real time data.
- 2) Gateways used to perform data analytics and aggregates real-time data from sensors
- 3) Communication infrastructure used to connect gateways.
- 4) Servers' used to collect mine and warehouse the all data.

B. *Big Data*: The smart city communication develops large amount of heterogeneous data. To analyze those data we require some major aspects of Big Data.

1) *Big Data Management*: The smart city applications are generates large volumes of heterogeneous data in different formats. This data is collected in different amounts and on a regular basis. To make sure useful utilization of this data, it is important to have effective big data management tools. Big data management contains development and policies, procedures and execution of different architectures. As we know the data comes from variety of sources, there is a need for data management that will used to recognize, manage, classify in the different formats and sources. Data management also provides scalable maintenance for massive data [25], [26], [27].

2) *Big Data Platforms for processing*: Big data used to perform data analytics that requires high processing capabilities. For this we require high quality reliable softwares and hardwares. The platforms should offer high performance computing capabilities in optimized way that supports stream processing and fault tolerance capabilities. There are different platforms are available for Big Data Analytics such as Apache Mahout, SAMOA, HPCC, Hadoop Map Reduce, Stratospher and IBM Infosphere Streams [31], they provides the stream processing of real-time big data applications for smart cities. Big data can be also processed on the Cloud platform using Platform as a Service (PaaS) and Infrastructure as a Service (IaaS) [32].

3) *Smart network infrastructure*: Most big data applications for smart cities require having networks connecting their components like smart cars with GPS system, smart house devices with voice instructions, and smart phones. This network should be capable of efficiently transferring all the data from their sources to the server where data is stored, processed and collected to transfer responses back to the different components of the smart city. The distributed approach is more suitable for huge data. This is important to preserve the quality of the decision making process in the real-time big-data applications. [19].

4) *Advanced software and algorithms*: Big Data applications having their unique requirements and pressing that need for high volume high speed processing. Standard algorithms used in regular applications may not be efficient enough to process. These algorithms need to be optimized to handle high data volumes, large variety of data types, distributed components and time constraints on decision making processes, including various geographical locations. In addition, these algorithms need to work effectively across heterogeneous environments. The algorithm solutions must be capable and reliable of operating and managing in highly different dynamic environments [33].

5) *Security and Privacy*: As the most data collected and processed in smart city applications, that contains some form of private information, it is important to secure the all data at all levels. Basic security concerns are data privacy and data authentication.

6) *Citizen Awareness*: In the growing countries citizen must be smart and aware to use Internet of Things solutions. Their active use require in providing needed information related to the different aspects and issues they may encounter with smart city IoT applications. Awareness will help to enhance quality and performance of the smart city applications. [26].

7) *Government Role*: Governing entities of smart cities must provide solutions and guidance for illiterate citizens. They may provide the environment of openness and transparency. Citizen's active participation and collaboration to keep, flow and the exchange of big data under control with security and privacy [10].

III. KEYS AND APPROCHES

The following eight key activities can be identified that often define a smart city [13]

- Smart Governance
- Smart Infrastructure
- Smart Building
- Smart Connectivity
- Smart Healthcare
- Smart Energy

- Smart Mobility
- Smart Citizens Equations

A. *Five Layer Approach:*

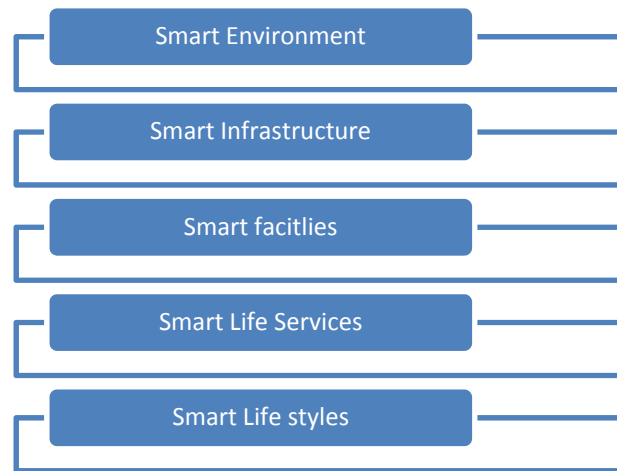


Figure 1: Proposed Layered solution of smart city development.

- 1) First Layer is Smart Environment; like smart real estate development in city.
- 2) Second layer is Smart Infrastructure like smart development of Rivers, Dams, Offices, and Roads etc. of a city.
- 3) Third Layer is Smart Facilities provide by government authorities in terms of E-banking, E-Lobby, Smart bank and smart government solutions.
- 4) Fourth layer contains Smart life services like e-hospital, smart living, use Internet of Things in basic needs.
- 5) Fifth layer is lifestyles, art and culture; this connects us with basics of ours.

These layers can be differentiating in different perspectives. Basically we are seeking the solutions for these layers in the terms of software's, Internet of Things, Data Storing and analytics etc.

IV. BENEFITS AND OPPORTUNITIES

As we know in India currently, many cities are competing to be a smart city that gives benefits environmentally, socially and economically. For the same, we are going to discuss some of the benefits, opportunities and approaches that may help to develop a city to a smart city. These solutions may be possible to achieve and enhanced different levels of resilience, governance and sustainability. In addition to improve the quality of life we need solutions for infrastructures, natural resources and management [2].

Some of the benefits of having a smart city include the following:

1. Efficient resource utilization
2. Better quality of life
3. Higher levels of transparency

V. CHALLENGES

Developers are facing many challenges in the design phase, development phase and deployment phase of big data applications for smart cities. The concepts behind smart cities are dynamic and elaborative environments, so it is important to develop distributed and flexible technologies. The whole requirements of smart cities are divided in different parts and all parts having their individual challenges.

The basic challenges with available big data tools are analytics, accessibility and cost. This will affect the performance of applications and services of smart city [8]. Some of the key challenges using big data are following.

Data sources and characteristics: Heterogeneous Data is generated from different sources in different variety. They can be unstructured in form like text, images, video, audio, social sites data etc. The developer needs to manage and classify unstructured data format into structured data format [7]. As we know big data identified with different Vs, where basic 3 Vs are: Velocity, Volume and Variety. We can include more on it on the basis of data requirements they are Validity, Veracity, Volatility, Value [2] and Variability [20]. The current algorithms, methods and software tools of data mining cannot handle the size and complexity of data. Developers are facing more challenges like data analytics, system architecture, data evaluation, big data distributed mining, maximum time evolving data, data compression, data visualization, and hidden data [8].

Data and information Sharing: The second most important challenge for developers is information sharing among distributed and heterogeneous users of different departments of smart cities. As we know all the government authorities having their own warehouse or agency that contains private, sensitive and confidential information. At the security level the sensitive data sharing is a big challenge. In this era all users are containing their private information on smart phones. The data of personal information of all the citizens, like personal ID's, addresses, numbers etc. are very sensitive and cannot be shared with all. The governments have to ensure that the privacy of citizens cannot be share [20].

Data Quality: The smart city will generate large amount of data, developers need to keep only important and informative data. As to store large amount of data distributed databases is needed, to balance the server load developers need to store only quality data in standard formats [21]. The challenge for developer is to keep data consistency, integrity and heterogeneity at large level otherwise that will cause challenges like data uncertainty and timeliness.

Security and privacy: other major challenge in a smart city using big data is security and privacy issues as we discussed in previous point, the sensitive information can't be share. In basic, developer needs to apply high level security and privacy algorithms on confidential information of government and citizens, and protect data from malicious attacks and unauthorized use. In addition, applications are integrated together across different networks, which require high level security when the data move over different types of networks [20].

Cost: The specifically very elementary challenges are to understand the meaningful data from the junk data that anticipated into the internet, refer as "Smart Data". Eighty-five percent of the entire data are noisy or meaningless. It is a very tough often assigned to verify and separate to refine the data from the noisy junk. In this era, all the organizations compromise with the already defined methods of conventional data. As latest technology, cloud computing and virtualization complicating the strength of mind or will to give a large amount of data solutions.

However, those technologies is gives the higher cost that all the organizations have to compromise. As we know, Data is growing up very quickly that it is enhancing expensive way to manage, store and analyze. The organizations hard try to find out how extended in space or time, where some of the data is meaningful, while other data are meaningless having no advantage hardly any hours after it have existed. To build Big Data solutions we will require new tools & technologies [2] [4] [5].

In smart city development high level software and hardware required, For example: sensors, Internet of Things, distributed databases etc. Cost is a sensitive issue because it is a government system developed for citizens. Government need to create a cost effective solutions. For Example an Energy usage reduction system [11] contains new components, features and systems government has to implement this model however, it is a very expensive to implement [16].

VI. CONCLUSION

Smart city and big data are two new and important concepts. Both contain same properties and challenges. To develop smart city applications we will get Big Data problems. This study gives both concepts and their definitions, challenges and some common attribute.

VII. REFERENCES

- [1] 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] Kitchen R. The real-time city? Big data and smart urbanism. *GeoJournal*. 2014;79(1):1–14.
- [4] Townsend AM 2013. Smart cities: big data, civic hackers, and the quest for a new utopia. WW Norton & Company.
- [5] Batty M. Big data, smart cities and city planning. *Dialogues Hum Geog*. 2013;3(3):274–9.
- [6] Vilajosana I, Llosa J, Martinez B, Domingo-Prieto M, Angles A, Vilajosana X. Bootstrapping smart cities through a self-sustainable model based on big data flows. *Commun Mag, IEEE*. 2013;51(6):128–34.
- [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] Fan W, Bifet A. Mining big data: current status, and forecast to the future. *ACM SIGKDD Explor Newsl*. 2013;14(2):1–5.
- [9] Al-Hader M, Rodzi A. The smart city infrastructure development & monitoring. *Theor Empir Res Urban Manage*. 2009;4(2):87–94.
- [10] 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.
- [11] 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.
- [12] Neirotti P, De Marco A, Cagliano AC, Mangano G, Scorrano F. Current trends in Smart City initiatives: Some stylised facts. *Cities*. 2014;38:25–36.
- [13] Tantatsanawong P, Kawtrakul A, Lertwipatrakul W. Enabling future education with smart services. In *SRII Global Conference (SRII)*, 2011 Annual IEEE; 2011. pp. 550–556.
- [14] West DM. Big Data for Education: Data Mining, Data Analytics, and Web Dashboards. *Governance Studies at Brookings*. 2012.
- [15] Marsh O, Maurov-Horvat L, Stevenson O. Big Data and Education: What's the Big Idea?. *UCL Policy Briefing*. 2014.
- [16] Aguilera G, Galan JL, Campos JC, Rodriguez P. An Accelerated-Time Simulation for Traffic Flow in a Smart City. *FEMTEC*. 2013;2013:26.
- [17] U.S. Department of Energy, "Smart Grid / Department of Energy,"

- [18] Yin J, Sharma P, Gorton I, Akyoli, B. Large-Scale Data Challenges in Future Power Grids. In Service Oriented System Engineering (SOSE), 2013 IEEE 7th International Symposium on IEEE; 2013. pp. 324–328.
- [19] 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.
- [20] Khan M, Uddin MF, Gupta N. Seven V’s of Big Data understanding Big Data to extract value. In American Society for Engineering Education (ASEE Zone 1), 2014 Zone 1 Conference of the IEEE; 2014. pp. 1–5.
- [21] 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.
- [22] 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.
- [23] Kim GH, Trimi S, Chung JH. Big-data applications in the government sector. *Commun ACM*. 2014;57(3):78–85.
- [24] Chourabi H, Nam T, Walker S, Gil-Garcia JR, Mellouli S, Nahon K, Scholl HJ. Understanding smart cities: An integrative framework. In System Science (HICSS), 2012 45th Hawaii International Conference on IEEE; 2012. pp. 2289–2297.
- [25] Xiaofeng M, Xiang C. Big data management: concepts, techniques and challenges [J]. *J Comput Res Dev*. 2013;1:98.
- [26] Borkar V, Carey MJ, Li C. Inside Big Data management: ogres, onions, or parfais?. In Proceedings of the 15th International Conference on Extending Database Technology. ACM; 2012. pp. 3–14.
- [27] Chaudhuri S. What next?: a half-dozen data management research goals for big data and the cloud. In Proceedings of the 31st symposium on Principles of Database Systems. ACM; 2012. pp. 1–4.
- [28] Dittrich J, Quiané-Ruiz JA. Efficient big data processing in Hadoop MapReduce. *Proc VLDB Endowment*. 2012;5(12):2014–5.
- [29] Middleton A, Solutions PDLR. Hpc systems: Introduction to hpc (high-performance computing cluster). White paper, LexisNexis Risk Solutions; 2011.
- [30] Alexandrov A, Bergmann R, Ewen S, Freytag JC, Hueske F, Heise A, et al. The Stratosphere platform for big data analytics. *VLDB J*. 2014;23(6):939–64.
- [31] Biem A, Bouillet E, Feng H, Ranganathan A, Riabov A, Verscheure O, Moran C. Ibminfosphere streams for scalable, real-time, intelligent transportation services. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data ACM; 2010. pp. 1093–1104.
- [32] Ji C, Li Y, Qiu W, Awada U, Li K. Big data processing in cloud computing environments. In Pervasive Systems, Algorithms and Networks (ISPAN), 2012 12th International Symposium on IEEE; 2012. pp. 17–23.
- [33] Wu X, Zhu X, Wu GQ, Ding W. Data mining with big data. *IEEE Trans Knowl Data Eng*. 2014;26(1):97–107.