

A REVIEW ON BIG DATA ANALYTICS AND ITS ROLE IN PUBLIC SECTOR

Kapil Dagar¹ and Sunita Dhingra²

1. M.Tech Scholar, Department of Computer Science and Engineering, University Institute of Engineering and Technology, Mahrishi Dayanand University, Rohtak, Haryana, 124001
2. Assistant Professor, Department of Computer Science and Engineering, University Institute of Engineering and Technology, Mahrishi Dayanand University, Rohtak, Haryana, 124001

Abstract: The Information and communication technology is transforming lives day by day. With this there is also an exponential rise in the amount of data and information available. The buzz word 'Big data' has spread everywhere. It's potential also created curiosity in the policy makers who are struggling with problems like growing population and diverse opinions. This paper will try to define what Big Data is, how it come to popularity and what will be its role in shaping the future. It will also help in specifically analyzing the role of big data in public administration and in public sector, how big data analytics can be feasibly utilized in public sector. It will discuss about the opportunities and challenges before big data and its implementation in smart administration.

Keywords: Big Data, Public Sector, Smart Administration, Public administration

1. INTRODUCTION

The 21st century has witnessed very radical growth in the use of Information and communication technology. New technologies are emerging which are transforming the way of interaction with data and information. Technologies like cloud computing; IOT (Internet of things) has found a stable place in our digital life. The information generated by plethora of devices is very high and it has transcended beyond our previous standards of measuring data. In the 1990's the data was in the amount of Kilobytes and Megabytes but in the first two decades of 21st century the data generated is exponentially increased and reached the mark from Gigabytes to Petabytes, Yottabytes and Exabyte's.

This is where the concept of Big Data is introduced, when the data sets are grown so enormous that traditional database system cannot handle them efficiently. Big Data tools helps to gather, store, manage and process this data very efficiently in a simple manner. This ability of Big Data to interact with data make it suitable to fit around many industries where large volume of data is generated, so field where data is generated rapidly and collected at a tremendous rate have reference to big data.

Now a day's people are documenting all their life events online, they are making video's, they are doing transactions online, they are playing games and also they are using social media which is generating data that cannot be handled by traditional databases. In a single day billions of posts and tweets are posted on social sites such as Twitter and Facebook. And millions of IOT sensor working in collaboration with each other produce data that transcends beyond our capacity to interpret it properly. Traditional databases are not capable of handling such large amount of data [1]. This type of data is being categorized as big data.

Traditional databases only support relational database's which are handled by relational database management system (RDBMS). But there are some problems with RDBMS which can be overcome by use of Big Data. While RDBMS only capable for supporting structured data, Big Data has capabilities that it can support both structured and unstructured data. This make it eligible in most of the areas where unstructured data is primarily generated.

1.1 Big Data Analytics:

It uses the potential data to look for unknown patterns in it or any correlation among entities that can be helpful in gaining new insights about real world phenomenon's and making a base for taking decisions that are statistically correct [2]. Analytics make it feasible to derive intelligence from unstructured and structured data sets. It helps in better utilization of resources and saves time and money. The Big Data analytics insights is the main driver of growth in today's competitive world. Data analytics set a new field of data-intensive science where data is the primary source of research, it will completely change the way of research and science [3].

1.2 Characteristics of Big Data:

Generally big data is characterized by three main features popularly known as 3V's, the first one represents Volume, the second one represents Variety and last one represents Velocity. Apart from these three main features two extra features also have been added later making a 5V approach, the fourth one represents Value and the later one represents Veracity [4].

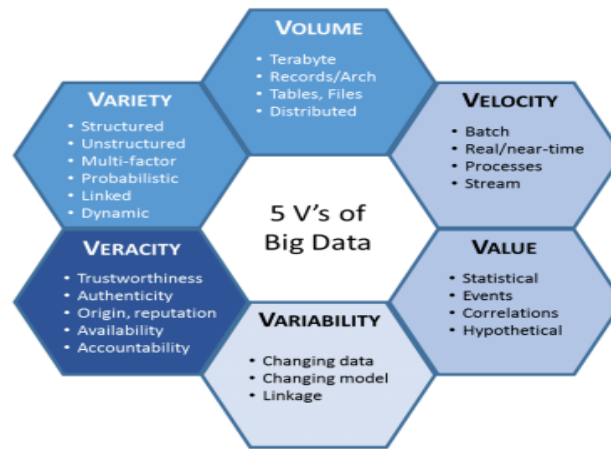


Fig 1.1 characteristics of big data [1.1]

- **Volume:** It is the quantity of all the data obtained from many sources. It is the data generated by all the users or all the connected devices collected at one place. It's very important to store all this data properly without any loss.
- **Variety:** It is the diversity present in the data i.e. different types of data collected in the form of text, audio, video or sensor data. The data may be structured form or in semi-structured form or completely unstructured form.
- **Velocity:** It is the speed at which data is generated day by day, the data transfer speed is valued very much in it because it must be accompanied by an infrastructure that can handle and provide optimal performance for the given velocity [5].
- **Value:** The term refers to the utility of the information that is being processing, it must provide certain values and should not to be completely useless in its implementation.
- **Veracity:** Veracity is very important feature in today's scenarios where lot of fake information is generated so it becomes very important for data to have correctness and accuracy. It should be free from things such as ambiguity, duplication and incompleteness.

1.3 Techniques and Technology Used

Big data uses various techniques and methods for its processing, some of the most popularly used techniques are Association rule learning, statistical analysis, correlation analysis, Cluster Analysis, Crowd Sourcing, Machine learning, Text analysis. The most popular software used for Big Data is Hadoop with MapReduce function which provide a complete ecosystem for processing Big Data combined with other tools such as YARN, Hbase, Zookeeper, Hive, Oozie, Mahout etc. It popularly uses HDFS (Hadoop distributed file system) for reliable data storage and MapReduce framework is used for processing data [6].

1.4 Methodology of Research

The research is conducted online by selecting articles and analyzing their content in the field of big data and public administration. Only relevant articles are filtered on a predefined set of criteria. The various sources of literature that were reviewed to conduct research includes books on big data, conference reports and thesis's, scientific research articles and journals. It also includes editorial articles and case studies.

Esteemed online journals such as Springer Link, Sage Publications and various other reputed journals found on the internet were also examined. The search is being done on Google Scholar where the following keywords were used "big data", "public sector big data" and "smart administration".

2. Role of Big Data in Public Sector

In the Republic of India, which witnessed a wave of digitalization when government bans the higher currency notes in 2016, more and more people started using digital form of transactions which generated huge data sets every day and night.

These type of data sets can be used to gain insights that can be very useful for future policy [7] The concept of smart cities is also around corner which will improve the quality of life for its citizens [8]. These insights are not limited to certain fields but can be applied to a variety of fields such as defense and internal security, energy distribution, education and research, health management, crowd management [9].

The population is grown very high and now it is very crucial for governments to gain insights from the big data and administer the needs of its people. With increasing smartphones in the hand of common people and the huge data generated by these, this data can be used to better administer the people by integrating it with big data analytics. The electronic devices are increasing so the data generated by them. As stated by a report by IDC (International Data Corporation), in the second quarter of 2018 the total number of smartphone shipped by all companies in India is 33.5 million units, with a growth of approximately 20% in shipment. [10] And one another prediction estimated that the total number of smartphone users present in India is to increase rapidly and will likely touch the mark of 442.5 million in 2022 [11].

Big Data can provide powerful knowledge that can be used for effective public administration by using optimal resources [12]. Simple big data solutions can be very effective for daily civil life problems such as in the field of Medical, Transport, Education etc [12] So it is very effective how public sector information can be used to serve some beneficial public interests.

Governments from a long time ago have been worked on creating knowledge and managing the existing knowledge that successfully helped them in a way to better governance [13]. The new form of technology will be completely citizen centered because the main components of government will be its citizens and they are the primary data generators who are also the primary feedback givers [14]. Government can also work on public policy that make the publicly available data open for all to increase the level of transparency in the governance and its easy accessibility for all for a better decision-making. As a democratic directive transparency and accessibility are fundamental to public policy making [15].

Many governments in the world are establishing data intensive research centers, which will help to gain valuable insights by manipulating large masses of information. Such an example is the “Big Data Research and Development Initiative” by Obama administration and Australian Government’s Information Management Office (2013).

The government in Albania also started a e-portal for its more than 571 e-services and open access to public data which showed a very positive feedback [16]. Other governments around the world are trying to create “open data platforms” such that all the valuable information to the community remain available to everyone and support better decision making. It should not result in any kind of ‘Digital-divide’. Thus open data platforms will result in different ways of dealing with acquired data, a variety of conclusions derived from it and plenty of ideas where this data can be employed in public sector.

When dealing with public sector data, it has to deal with multiple departments, they generate very high volumes on data on a daily basis and they spent huge resources on collecting this data and properly storing it. The primary goal after that is to make an infrastructure that can utilize these data resources in a more efficient way [17]. There must be collaboration between departments for sharing the data.

3. Applications of Big Data in Public Sector

Big Data can be used functionally in three aspects, the first of them is Public supervision i.e to find irregularities or taking responsive action to the events happen in society, the second aspect is public regulation i.e compliance of constitutional rules and necessary regulation in society and the third aspect is public service delivery i.e. to maximize the distribution and reach of the resources [18].

Its practical applications include multitude of fields such as Medical, Energy, Defense, Education, Transportations, Environment etc.

Some example of real life scenario for big data decision making are given -

- **Entity analytics:** An entity can be anything or any person, it can be analyzed for certain parameters and can be looked for correlations and pattern among them, that can be very helpful in future for predictive decision making and better policy making. For example, Different projects on Twitter have been launched to analyze and examine tweets posted online to understand the underlying emotions of people, decisions of people and their socio-economic condition of people which can be very helpful in knowing the opinions and trends [19].
- **Regulations:** Big data can be used for compliance of regulations in various fields like resource management, environment and natural resources, policy making etc.
- **Law Enforcement:** Accurate predictions can be made using big data analytics in law enforcement to prevent crimes. Big data prediction can drastically decrease the rate of crime by giving the insights that uncover underlying causes and motivation for crime, this is done by the process of examining a large set of data.
- **Tax Fraud Prevention:** Big data management of transactions can help in gaining very rich insights about financial movements which help in gaining insights for better economic policy making and prevent economic frauds to help boost the economy of a nation [20].
- **Social Services:** Government can use big data for optimal implementation of various social security schemes and social welfare schemes which will help in distributing the resources to every citizen in a transparent manner.

3.1 Challenges before Big Data in Public Sector

Apart from some core big data management challenges which include efficient collection and retrieval of data, processing of huge amount of data, proper visualization of data, proper interpretation of the collected data and quality assurance of the data, big data in public sector also face some distinctive challenges which have to be addressed [21].

Privacy: The first issue comes with big data in smart administration is the data privacy and implementation of transparency. Privacy is very crucial for any citizen and it is an important feature of democracy and if big data collection violate the fundamental of democracy then it will have to face strict prohibitions, so it become very important factor for policy makers to implement data policy in a way such that it cannot violate personal spaces of people while at the same time have a robust data infrastructure. There must be transparency and control over how the data will be collected, used and managed.

Security: The second challenge before big data in public sector is security of the data, take an example of healthcare, the health data is very confidential and sensitive and its security can't be compromised, and it is also applicable for data in the field of security and energy. So to build a secure infrastructure around big data is also a primary goal of the policy makers.

Implementation: The third challenge before big data is its practical implementation. It has to identify key areas where the derived insights and knowledge will foster growth and provide value rather than predictions that cannot be practically useful. It must provide with mechanism that can be useful and efficient for public services.

Table 1. findings of different authors.

S.no.	Title	Author/Year	Findings/Contributions
[1]	Taming the big data tidal wave	Franks (2012)	Traditional database are not capable of handling very large amount of data.
[2]	Big data: A revolution that will transform how we live, work and think.	Mayer-Schonberger & Cukier (2013)	Big data will help in gaining new insights about real world phenomenon's that will completely change our view on data.
[3]	Critical questions for big data. Information, Communication & Society	Boyd & Crawford (2012)	Data science will completely change our thinking of science and research.
[4]	How 'big data' can make big impact	Fosso Wamba (2015)	The 5V model for big data characteristics.
[5]	Big data issues and challenges moving forward	Kaisler et al. (2013)	Defines how velocity of the data generated, and speed of data transfer are crucial in management of data.
[6]	A Review paper on Big Data: Technologies, Tools and Trends	Anurag Agrahari (2017)	Various tools and techniques used in core big data framework.
[7]	Accelerating democracy: Transforming governance through technology	John O. McGinnis (2012)	How growing datasets can be used to gain insights in e governance
[8]	On the Internet of Things, smart cities and the WHO Healthy Cities	Kamel Boulos & Al-Shorbaji (2014)	Smart cities and their impact on quality of life.
[9]	Agencies rally to tackle Big Data	Mervis (2012)	Big data derived insights can be applied to a multitude of fields and not limited to a few.
[10]	Smartphone usage in India	Report by IDC (2018)	In second quarter of 2018 total number of smartphone shipped by all companies in India is 33.5 million units a whopping 20% growth
[11]	Prediction on smartphone usage in India	Economic Prediction (2018)	Indians will be using 442.5 million smartphones by 2022. Huge amount of Data will be generated.
[12]	The Value and Challenges of Public Sector Information, Cosmopolitan Civil Societies	Henninger (2013)	Effective public administration and optimal usage of public resources using Big Data In various fields of Medical, Transport, Education etc.
[13]	Beneficence and the Expert Bureaucracy	Jordan & Sara R, (2014)	Governments from a long time ago have been worked in field of creating knowledge and managing the knowledge
[14]	Survey on Big Data Analytics in Public Sector of Russian Federation	Anna et al. (2015)	Importance of citizen feedback in implementing Big Data technology
[15]	On making public policy with publicly available data	Napoli & Karaganis (2010)	Publicly available data open for all to increase the level of transparency in the governance and its easy accessibility for all for a better decision-making
[16]	Big Data in e-Government Environments: Albania as a Case Study	Elezaj et al. (2018)	Big Data case study of Albania government, its implementation and feedback.
[17]	Big Data in the Public Sector: Lessons for Practitioners and Scholars. Administration & Society	Desouza et al. (2014)	How different departments can collaborate on sharing data for effective governance.
[18]	To do more, better, faster and more cheaply: using big data in public administration	Maciejewski (2017)	Different functions of big data in public administration.
[19]	Discovering global socio economic trends hidden in big data.	Lopez & Amand, 2012	Twitter project to analyze emotions, decisions and socio-economic conditions of people using Big data.
[20]	New Horizons for a Data-Driven Economy	Munné R. (2016)	How can big data help economy in a dramatic way.

[21]	Big Data in the Public Sector: A Systematic Literature Review	Zhan & Ming (2017)	Challenges in big data implementation and effective working.
------	---	--------------------	--

4. CONCLUSION

The aim to conduct this literature review is to get familiar with big data and gain a comprehensive understanding of public administration using big data. It is evident from growing no. of electronic devices that in the future our technological infrastructure will face enormous amount of data. Big data with its technology such as Hadoop, can effectively be used to acquire, store and process this data successfully which is not possible with traditional tools. Big data can be combined with other technologies like cloud computing and internet of things to provide us with real time insights and valuable knowledge. Although the big data has been successfully deployed by some companies in private sector, its role in public sector is still undermined. There are many sectors where big data can be used for daily civil life problems to gain effective solutions. But there are some challenges also associated with big data in public administration such as how it will deal with privacy and security and how to achieve transparency. The research in Big data in public sector is very crucial in context of growing population and its effective management. The insights that big data can provide could be very important for our policy making. There is still a huge scope of research in this area.

References

- [1] Franks, B. (2012). "Taming the big data tidal wave", Hoboken, NJ: John Wiley, 1-27, DOI: 10.1002/9781119204275
- [2] Mayer-Schönberger, V., & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Boston, MA, Houghton Mifflin Harcourt, 21-35, DOI: 10.1080/15228053.2016.1220197.
- [3] Boyd, D., & Crawford, K. (2012), "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon" *Information, Communication & Society*, 15(5): 662-679, DOI: <http://dx.doi.org/10.1080/1369118X.2012.678878>
- [4] Fosso Wamba, Samuel & Akter, Shahriar & Edwards, Andrew & Chopin, Geoffrey & Gnanzou, Denis (2015) "How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study", *International Journal of Production Economics*, 165, 234-246, DOI: <https://doi.org/10.1016/j.ijpe.2014.12.031>
- [5] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013) "Big data issues and challenges moving forward", *Proceedings from the 46th Hawaii International Conference on System Sciences*, Wailea, Maui, HI, 995-1004, DOI: 10.1109/HICSS.2013.645
- [6] Anurag Agrahari, Prof D.T.V. Dharmaji Rao (2017) "A Review paper on Big Data: Technologies, Tools and Trends", *International Research Journal of Engineering and Technology (IRJET)*, 4(10): 640-649
- [7] McGinnis, J. O. (2012) "Accelerating democracy: Transforming governance through technology", Princeton, NJ: Princeton University Press, 1-40, DOI:10.1111/1478-9302.12073_118
- [8] Kamel Boulos, Maged N. & Al-Shorbaji, Najeeb M. (2014) "On the Internet of Things, smart cities and the WHO Healthy Cities", *International Journal of Health Geographics*, 13(1): 1-6, DOI: 10.1186/1476-072X-13-10
- [9] Mervis, J. (2012) "Agencies rally to tackle Big Data", *Science (New York, N.Y.)*, 336, 22. DOI: 10.1126/science.336.6077.22.
- [10] Retrieved From <https://www.idc.com/getdoc.jsp?containerId=prAP44210918> [accessed 23 Nov, 2018]
- [11] Retrieved From <https://www.statista.com/topics/4600/smartphone-market-in-india/> [accessed 23 Nov, 2018]
- [12] Henninger, Maureen (2013) "The Value and Challenges of Public Sector Information", *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, 5(3): 75-95. DOI: <https://doi.org/10.5130/ccs.v5i3.3429>
- [13] Jordan, Sara R. (2014) "Beneficence and the Expert Bureaucracy", *Public Integrity* 16(4): 375-394. DOI: 10.2753/PIN1099-9922160404.
- [14] Anna, Kuraeva & Nikolay, Kazantsev (2015) "Survey on Big Data Analytics in Public Sector of Russian Federation", 3rd International Conference on Information Technology and Quantitative Management, ITQM 2015, 55: 905-911. DOI: 10.1016/j.procs.2015.07.144.
- [15] Napoli, Philip M. & Karaganis, Joe (2010) "On making public policy with publicly available data: The case of U.S. communications policymaking", *Government Information Quarterly*, 27(4): 384-391 DOI: 10.1016/j.giq.2010.06.005
- [16] Elezaj, Ogerta & Tole, Dhimiter & Baci, Nevila. (2018) "Big Data in e-Government Environments: Albania as a Case Study", *Academic Journal of Interdisciplinary Studies*. 7(2): 117-124, DOI: 10.2478/ajis-2018-0052.
- [17] Desouza, Kevin & Jacob, B. (2014). "Big Data in the Public Sector: Lessons for Practitioners and Scholars" *Administration & Society*, 49(7): 1043-1064, DOI: 10.1177/0095399714555751.
- [18] Maciejewski, M. (2017) "To do more, better, faster and more cheaply: using big data in public administration", *International Review of Administrative Sciences*, 83: 120-135, DOI: [10.1177/0020852316640058](https://doi.org/10.1177/0020852316640058)

[19] Lopez, G., & Amand, W. S. (2012). "Discovering global socio economic trends hidden in big data", Retrieved from <https://www.unglobalpulse.org/discoveringtrendsinbigdata-CHguestpost> [accessed 27 Nov, 2018]

[20] Munné, R.: Big data in the public sector. In: Cavanillas, J.M., Curry, E., Wahlster, W. (eds.) "New Horizons for a Data-Driven Economy", Springer, Cham (2016). 195–208, DOI: [10.1007/978-3-319-21569-3_11](https://doi.org/10.1007/978-3-319-21569-3_11) [Google Scholar](#)

[21] Zhan, Ming. (2017). "Big Data in the Public Sector: A Systematic Literature Review", Scandinavian Journal of Public Administration, 21(3): 39-61

[1.1] Exploring Techniques of Improving Security and Privacy in Big Data - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/The-5V-of-Big-Data-Characteristics_fig1_321050765 [accessed 23 Dec, 2018]

