Big Data In E-Governance Management

Harshita Saluja[1*], Pallavi Asthana[2], Sumita Mishra[3], Sachin Kumar[4], Bramah Hazela[5]

Corresponding Author: *harshitakaur666@gmail.com,* *Tel.: 9793403564*

Abstract— Large amount of data produced through humans, machines and automated system can be monitored for improving the efficiency of systems. This data can be real time monitored through Big Data analysis. Big Data could be structured, unstructured or semi-structured and is characterized by Volume, Velocity, Variety, Veracity and Value. The data is novel, dynamic, and scalable. Data analysis performed on Big Data can be highly useful in E-governance. This paper explores the five ways in which big data is characterized, methods of its classification, Big data management and its role in effective implementation of e-governance like education, health care, revenue etc., where huge amount of data is generated. This data is useful in understanding the factors that can be monitored and analyzed for improvement and betterment of current policies initiated by government.

Keywords— Big Data, Dynamic, Scslable, Data Silos, E-governance.

I. INTRODUCTION

Big Data is a stack or compilation of massive and huge amount of data. It is the data which has been generated from past, which is being generated in present and which will be generated in future. The data from past has grown to a large scale within a span of two decades and is growing every day, every second with an exponentially high rate. With the advancements in technology, a vast variety of data from various domains has been generated.

"Big data refers to the data or sets of records that are too large in volume to be operated using the existing database management tools and techniques."[1]

This refers to the idea that data is growing continuously at a fast rate. Facebook sets a record of login of 1 billion people per day. The increment in the volumetric rate of data is very high as compared to the growth of data in past decades. Every alternate day we create information that only comprised of the volume of data created in the beginning of 2003. The data is growing in such huge quantity that the total amount of data collected and stacked by industries double in every 1.2 years.

Google, on an average, alone processes over 40 thousand search queries per second for over 3.5 billion in a single day. This evolution of various social networking sites, organizations, and technologies result into big size of databases that cannot be handled by traditional databases. Data is being generated, transferred, and received in the capacity of terabytes to petabytes. Every minute, people send 204 million emails. People generate 1.8 million facebook links, send 278 thousand tweets, upload 200,000 photos

to facebook, and around 100 hours of videos are uploaded to youtube every minute[2,3]. It may be inferred that data science will act as catalyst for economic growth.
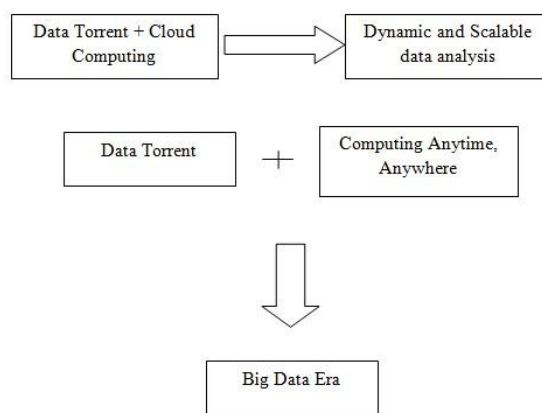

Fig 1: General Overview of Big Data Formalisation

II. CHARACTERISTICS & CLASSIFICATION OF BIG DATA

Big data is characterized as following [4,5]:

(i)Volume: The amount with which data is being created is defined under the first V, that is, volume. It represents that dimension of data which shows the huge amount of data which is generated. AT&T is considered to hold world's largest volume of data in one unique database; its phone records database is 312 terabytes in size, and contains almost 2 trillion rows.

(ii)Velocity: It is the rate with which the data is being generated. It has two aspects that can be described as data moving in pipes and latency that refers data in motion or measure of velocity. According to survey, 570+ new websites come into existence every minute of every day.

(iii)Variety: It represents the type of data or category of data like text, audio, video, pdfs, etc.

(iv)Veracity: It represents the interpretations caused by the change in the structure of data.

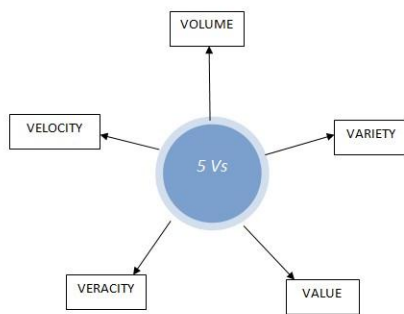**(v)Value:** It shows the value of the data we are dealing with in different aspects.

Fig 2: Characteristics of Big Data

Big Data can be classsified into three categories and other techniques that work on these data are derived on the basis of the category of data which is required to be processed [6,7].

(i) Unstructured Data: It is the most abundant form of data. It is the complex data which cannot be put into a particularly set pattern or format. For example: data that from social networking sites like facebook, linkedin, logs, web-chats, youtube. This data is textual, images and numeric also.
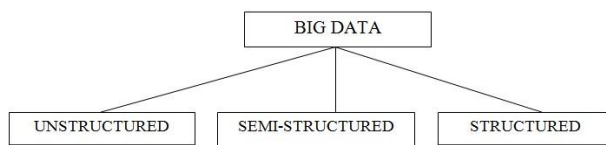

Fig 3: Classification of Big Data

(ii) Semi-Structured Data: This data cannot be categorized as perfect pattern data or without pattern, the data is neither structured nor unstructured. It is in mixed form. For instance, call centre calls with toll-free responses, web logs that track website activities etc.

(iii)Structured Data: This is the most planned and worked out data with proper categorizations, formats and set patterns. Organizations act as the source for structured data. For instance: enterprise resource planning, back up storage for large volumes of data in particular format of data sheets[8-9].

Now a days, vast amount of data is created that is dynamic and complex in nature. Main sources of big data are :

(i)Machines: Data generated from machine is the data generated from real time sensors in Industrial machines, sensors or logging of vehicles.

(ii)People: This data is generated by people in daily activities .It is most abundant and vast amount of data generated through social media, status updates, tweets, photos, audios, videos, media, etc.

(iii)Organizations: Organizational generated data is somewhat traditional types of data that includes structured form of data[10,12].

"Data exhaust" refers to ambient data that are passively collected, non-core data with limited or zero value to the original data-collection partner.

III.   MANAGEMENT OF BIG DATA

Data management is the most complicated issue in the field of big data in the present scenario . Traditional methods and techniques of data handling and management cannot be used with this tremendous amount of data as they are expensive, manual, and complex. They are not suitable for handling this huge quantity of data which is being generated every minute. Big data management that requires business values are: Customer loyalty, Operational efficiency, Fraud detection,Security risk and Predictive maintenance.

Big data management techniques provide self service i.e., scientists and organizations don't need to wait for fully certified data instead they can focus on insights and the other is operational agility i.e., engineers and architects can deliver actionable insights to business applications in real time. Big Data solutions can transform business processes as Procurement, Product Development, Manufacturing, Distribution Marketing, Pricing and Yield Management, Sales and Human Resources[13,15].

The data is being managed in data ware houses and data silos. Data silos are the huge stacks of data that has been generated through various sources. Data ware house is the data base where all of this data is stored.

IV.   ROLE OF BIG DATA IN E-GOVERNANCE

E-governance is about making all the services easy for all the people through Information and Communication Technology. This is important for reliable and transparent mechanism through multiple outlet to provide better and cost effective services .This process requires the maintenance of large amount of Data and Big Data can play the key role in generation , maintenance and analysis of this huge amount of Data. The analysis further provides the results that can play remarkable role in future decisions. So, techniques in Big data analytics are highly useful for further improvement in planning and implementation of e-governance projects. The benefits are achieved in terms of improvement in online information system (resource management), service delivery for the purpose of business analytics through open processeses(data organization and management) ,visualization of government performance(visualization report).and utilizing it for new business projects(decisions) and improving business growth[16-17].

Big Data Analytics paves a better way for e-governance in following ways:

(a) It can work on raw , structured and semi structured data where traditional method can only work on structured data.

(b) In the traditional practices, statics analytics can be done with known conditions where as in the environment of Big data various unknown business questions can also be explored.

(c) Traditional methods have its limitation in terms of the capacity e.g. data with which it can work remains within the terra-bytes , In big data large and messy data can be reasonably handled with the help of various tools and techniques.

(d) Big data analytics can work with graph, audio and video analytics and provide real time analysis of data instead of waiting for the completion of transformation jobs that are batch oriented and applicable for text analytics only.

E-Governance services are provided on the basis of four models: Government to Employees, Government to citizens, government to government and government to business. Various e-governance programs are as National e-governance plan(NeGP) formulated by Department of Electronics and Information Technology and Department of Administrative reforms and Public grievances[18] .

In agriculture, there are projects like AGMARKET and SEEDNET by Department of Marketing and Inspection by ministry of agriculture and Government of India. Other Services include Issue of Birth /Death certificates ,

Supply of water and sanitation in rural areas and provision of Unique Identification number Aadhar[19].

Areas and Applications where e-governance can benefit from big data analysis are summarized in the

Table 1 below:

Area	Applications	Type of decisions using analytics
Education	Student tracking, School location Planning, Administration	Finding various factors that contribute to student success. Identifying and planning location best suited for building school Planning and testing standards of schools
Health Care	Disease detection, Epidemiology, Disease prevention	Finding and reporting the occurrence of diseases. Identifying the causes, distribution and control of diseases. Identifying factors that can prevent disease occurrence
Public	Problem Analysis, Fulfilling Information, needs	Helps in analyzing problem faced by public and finding solutions to it. Identifying and providing information requirements of Citizens.
Safety	Crime Analysis, Record Keeping	Identifying type of crime and areas of occurrence Helps in keeping records of all crimes that have occurred in the past
Revenue	Tax assessment, Fraud Analysis	Calculating tax assessments and checking whether they are correct Analyzing the fraud occurrence
Environment	Environment analysis, Water/Air quality testing	Understanding which factors contribute to healthy ecosystem Ensuring water/air standards are met.
Business	Cost/benefit analysis, Employee satisfaction, Resource planning	Understanding what can lead to be more cost effective. Finding employee expectations, Identifying required resources

Big Data analysis is proven as a useful technique for e-governance projects, but still, it has some challenges in the complete implementation in terms of Privacy, Security, Shortage of technical and qualifies people to handle and implement such techniques, lack of relevant software to integrate multiple data sources and formats and mainly lack of interoperability due to the existence of traditional practices.

V. CONCLUSION

Big data analytic tools are able to process and store the data generated in daily lives and that too in bulk. With the advancements in technology, the amount of data has also grown exponentially that too at high rate. New technologies have ability to record, monitor measure and merge all types of data. So, industries need new ways, tools and technologies for tracking, analyzing and managing data. Big Data can provide tremendous opportunities in the area of e-governance by utilising it efficiently and in a secure way. This is a area where large data in stored and the it needs to converted in to the implementation of the policies that are beneficial for the people of the country.

REFERENCES

[1] Big Data: Characteristics, Challenges and Data Mining, International Journal of Computer Applications (0975-8887); International Conference on Advances in Information Technology and Management ICAIM-2016s.
[2] Introduction to Big Data, Ilkay Altinas, www.coursera.com.
[3] Hiba Jasim Hadi, Ammar Hameed Shnain, Sarah Hadishaheed, Azizahbt Haji Ahmad,' BIG DATA AND FIVE V'S CHARACTERISTICS', International Journal of Advances in Electronics and Computer Science, ISSN: 2393-2835, Volume-2, Issue-1, Jan.-2015
[4] Walunj Swapnil K., Yadav Anil H., Sonu Gupta,' Big Data: Characteristics, Challenges and Data Mining', International Journal of Computer Applications (0975 – 8887)-2016.
[5] Motashim Rasool, Wasim Khan, Big Data: Study in Structured and Unstructured Data, International Journal of Technology Innovations and Research (IJTIR).
[6] K.Arun Dr.L.Jabasheela 'Big Data: Review, Classification and Analysis Survey', International Journal of Innovative Research in Information Security, ISSN: 2349-7017 Volume 1 Issue September 2014.
[7] Praful Koturwar,SheetalGirase, Debajyoti Mukhopadhyay 'A Survey of Classification Techniques in the Area of Big Data',
[8] YongjunPiao, Hyun Woo Park, Cheng Hao Jin, Keun Ho Ryu, "Ensemble Method for Classification of High-Dimensional Data," 978-1-4799-3919-0/14, 2014, IEEE.
[9] Sindhujaa N , Vanitha C N , Subaira A S ,' An Improved Version of Big Data Classification and Clustering using Graph Search Technique', International Journal of Computer Science and Mobile Computing, Vol. 5, Issue.2, February 2016, pg.224 – 229.
[10] Managing Big Data. An interview with David Gorbet ODBMS Industry WatchJuly2012http://www.odbms.org/blog/2012/07/managin 2, g-big-data-an interview-with-david-gorbet
[11] Bakshi, K., 2012. Considerations for big data: Architecture and approach. In: 2012 IEEE Aerospace Conference, Big SkyMontana. pp. 1-7.
[12] Agneeswaran, V., 2012. Big-data - Theoretical, engineering and analytics perspective. In: Srinivasa, S., Bhatnagar, V. (Eds.), Big Data Analytics. Vol. 7678 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 8-15.
[13] Andrzej Chluski, Leszek Ziora 'The role of Big Data Solutions in the management of organizations, review of selected practical examples', International Conference on Communication, Management and Information Technology (ICCMIT, 2015).
[14] Schmarzo B.Big Data: Understanding how data powers businesses', Wiley 2013.
[15] Gerard George, Martine R. Haas,Alex Pentland, 'Big Data and Management', Academy of Management Journal 2014, Vol. 57.
[16] Preet Navdeep, Dr. Manish Arora, Dr. Neeraj Sharma,' Role of Big Data Analytics in Analyzing e-Governance Projects' GIAN JYOTI E-JOURNAL, Volume 6, Issue 2 (Apr-Jun 2016) ISSN 250-348X
[17] Rajagopalan M.R, Solaimurugan vellaipandiyan,' Big Data Framework for National e-Governance Plan' , Eleventh International Conference on ICT and Knowledge Engineering-2013.
[18] Department of Electronics and informstion technologies India, http://deity.gov.in/content/national-e- governanceplan.
[19] Aapka Aaadhaar, Unique identity & address in India. http://uidai.gov.in/.

Authors Profile

Ms. Harshita Kaur Saluja is the final year student of Bachelor of Technology. She is persuing her B.Tech in Electronics and Communication from Amity University. She has done her trainings on VLSI & Embedded Systems from Motilal Nehru National Institute of Technology, Allahabad and on Communication, Navigation, and Surveillance from Airport Traffic Control, Chaudhary Charan Singh Airport, Lucknow. She has also done certification course on Core Java. She has also presented paper on 'AI: Boon or Bane' in 2017 and 'Biometric Sensors' in 2016 in National workshops conducted in the University. Her main areas of interests include VLSI, Big Data, and CRC Algorithms and implementation on FPGA.

Ms.Pallavi Asthana is an Assistant Professor in Department of Electronics & Communication Engineering in Amity University. She is currently pursuing PhD. Her area of interests include Engineering Education, Machine Learning and Experiential Learning.

Dr.Sumita Mishra is currently the Assistant Professor in the Department of Electronics and Communication at Amity University, Lucknow. She has received Ph.D degree in

Electronics from DRML Awadh University, India and Masters of Technology in Optical Communication from SGSITS, Indore, India. She has published more than 20 referred publications. She is a member of IEEE, IET (U.K.), ACM, PCPro, IAENG, IACSIT. Her current research interests include Deep Learning, Visible Light Communication and Machine Vision.

Dr. Sachin Kumar has persued Bachelors of Engineering and Masters of Technology in Electronics and Communication. He has done Ph.D in Electronics. He is the member of IEEE. IETE, IAENG and IET(U.K). His research area include applications of Digital Image Processing, Bio-Medical Signal Processing, Machine Learning and Applications of Wireless Sensor Networks. He has an experience of more than 20 papers in various reputed Journals and has presented more than 40 papers in various reputed National and International Conferences. He has also worked as a member of Editorial Board/Reviewer in various journals.

Dr. Bramah Hazela is an Assistant Professor in the Department of Computer Science and Engineering at Amity University, Lucknow. His area of research is UML, Artificial Intelligence and Software Engineering.