



Big Data Analytics in E-Governance: A Review

Shantanu¹, Prashant Vibhor Agarwal², Prashant Dev³, Vijay Kumar Singh⁴, Gunjan Goswami⁵

¹ Computer Science & Engineering Department, JSSATE, Noida

² Computer Science & Engineering Department, JSSATE, Noida

³ Computer Science & Engineering Department, JSSATE, Noida

⁴ Computer Science & Engineering Department, JSSATE, Noida

⁵ Assistant Professor, Computer Science & Engineering Department, JSSATE, Noida

Abstract — Big data analytics play a pivotal role in the effective and efficient management of E-Governance. The amount of data generated and the pace, at which it is being accumulated, is rising exponentially. The proliferation of artificial intelligence, Web analytics, Internet of things and other types of emerging technologies are creating new data streams that only adds to traditional data stores, which gives rise to a new term called “Big Data”. Big data analytics is the process of uncovering the hidden patterns and mining knowledge in the sets of Big Data.

E-Governance in the modern world will lead to the creation of large data sets within countries like India where the government is planning to develop smart cities, which in future will be contributing a large chunk of data. There is a need to discover knowledge for E-Governance applications and to develop techniques to deal with the challenges of processing these large amounts of data and transforming them into a compatible homogeneous e-repository of useful information which can be directly used for analytical purposes. Thus this paper concentrates on the impact of big data analysis for the E-governance. Finally, this paper concludes with the conceptual knowledge of big data and its implementation in the public administration and E-Governance.

Keywords- Big Data, Public Sector, information and communication technologies (ICT), E-government, E-governance, Big Data Analytics

I. INTRODUCTION

With the rise information and communication technologies (ICT's), the production and consumption of data in the present day is very large and the numbers are evolving day by day. This humongous amount of data comes under the category of Big-Data. Big data is the new word which comes into existence during the rise of the 21st century with all technology spreading in every sector of life. So many data is being produced every single day but there is lack of tools to handle this much amount of data. So Big Data analytics is a technology that can help us to tackle the large data sets.

As we know the technology has evolved exponentially over the last decade and the rise of Smartphone's, satellite navigation, social networking, online banking, the stock market, smart vehicles, sensors are all contributing to data and more data.

Analyzing this data is also a problem which can be solved by various big data tools like Hadoop, Map Reduce, and Pig etcetera. This way government could become future proof to take on every problem.

Governments all over the world are also getting involved with the IT revolution happening in the world. Any government which uses the Information and Communication Technology (ICT) to operate the business and its own services are called an E-Government. They also face the same problem of all the data bombardment and floods of data, this is also a problem for governments of developed nation and developing nation. A government which has to make policies regarding the betterment of country has the huge responsibility to uncover the patterns of what are going, the trend and the need for people.

Generally, E-Governance makes up with these objectives in mind:

1. A Policy Framework.
2. Better Public Services.
3. Better and Cost Effective Government Operations.
4. Citizen Engagement.
5. Administrative and Institutional Reforms.

II. BIG DATA ANALYTICS

The word Big Data Analytics is an integration of two words 'Big Data' + 'Analytics'. Big data just means any data so large which can't be processed by traditional data processing technique. Large data started with the size of few Terabytes(1024 GB) at first but now at present ranging from some dozen Terabytes to few Petabytes (1024 TB) of data in a single dataset. Big data analytics help organizations to find the pattern and to capture the trend, which allows the organization to better forecast, more optimum resource management, better performance and increased customer base. Some of characteristics of Big Data are:

1. **Pocket-Friendly:** When it comes to storing large datasets big data technologies like Hadoop and Cloud-analytics are most cost effective.
2. **Better Forecasting:** The Hadoop provide tools and high analytic power which help the organization to make sense of all the data churning on the daily basis. Organizations are able to take decision much efficiently now.
3. **Scope for Improvement and Innovation:** With a tool that can predict and classify things in a way that no one other can. Organization always has a sight in future by which they can improve their products and help in producing new products too.

II.1 CHARACTERISTICS OF BIG DATA

Big Data is characterized by 4 V's i.e. volume, velocity, variety, veracity.

1. **Volume.** It defines the amount of data, i.e. how big data it is.
2. **Velocity.** It defines the rate at which updating of data takes place. Data update refers to the amount of new unprocessed data being generated at the real time or data added after. The velocity of data is in increasing state right now with each day new means of data are being added.
3. **Variety.** It refers to the different type of data that is being produced right now. Different type of medium produces a wide variety of data. Images, Videos, Excel Sheets, Simulation data, Sensor readings everything lead to a different type of data.
4. **Veracity.** It refers to the useless amount of data that won't contribute to information or knowledge building. Veracity could also be the incomplete datasets, noises present in datasets.

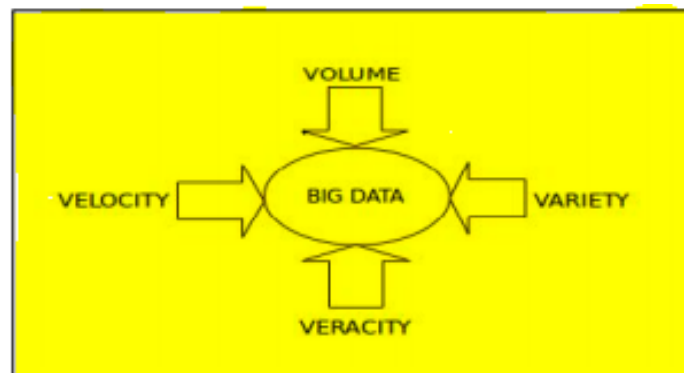


Figure 1. Characteristics of Big Data

III. E-GOVERNANCE MODELS

Recognizing the increasing importance of electronics, the Government of India established the Department of Electronics in 1970. And, they came up with the five principles of sincerity, contribution, liability, helpfulness and uniformity exercised within this E-governance Framework to support democratic governance. The concept of e-government started with the arrival of government websites in the 1990s. Before the improvement of e- governance, the situation of government was rigid, inactive hierarchical regulation while the web was energetic, stable and unfettered.

Union and State Governments had taken important e-governance initiatives in the last 10 to 15 years to assess their strength and weakness to improvise the democratic government system.

Government to citizens (G2C)

Government to employees (G2E)

Government to businessman (G2B)

Government to government (G2G) .

3.1. Government to Citizens (G2C)

In this model, varieties of ICT (Information and communications Technologies) services are offered to the citizens in an efficient and economical manner in order to increase the transparency between government and citizens using technology. Types of services which are included in this model are:-

- Transactions such as payment of services.
- Mundane services such as name or address changes.
- Copies of land-record.

3.2. Government to Businessmen (G2B)

It is the online cooperation between state and union government and the commercial business sector with the objective of providing businesses information and advice on business affairs with an objective to reduce difficulties for business provide immediate support.

They share information with this model like

- Collection of Revenue.
- Registration of patents.
- Payment of bills and penalty.
- Complaints or any kind of dissatisfaction

3.3. Government to Employees (G2E)

This model gives a platform such as online tools, sources, and articles that help employees to maintain a dialogue with the government and their own companies'. Some of the benefits of G2E model includes:

- E-payroll- to maintain the online sources for viewing paychecks, pay stubs, pay bills.
- E-benefits- allow employees to know about their benefits and rights.
- E-training- allows for fresh and existing employees to regularly maintain the training.

3.4. Government to Government (G2G)

In this model, most of the relevant information are shared between union and state government in order to improve the efficiency of both types of government. There is always a lot of information that is needed to be shared with various government agencies, department and organizations.

IV. BIG DATA ANALYTICS FOR E-GOVERNANCE

Big Data Analytics need for e governance is due to the following factors:-

Big Data, by definition, requires variety, implying various sources of information spread across department boundaries be available which requires an efficient architecture and effective infrastructure for data seizure, analytics, transfer, management. Making Data-driven decisions are still an art but also need analytics requiring trust in Data Scientists and swiftness in implementing decision. While Big Data technologies are production ready, Analytics is still experimental. Discovering patterns, knowledge, classification, predictive analysis and other insights from big data is a serious problem and provides a number of opportunities to innovate.

Data Analysts from sectors of eHealth, eEducation, eGovernance are yet to be identified so that the full potential of data is available. Deducing the most valuable insights from data requires intimate

Knowledge of the specific business areas. Determining what questions to ask and interpreting the results, combining Big Data Platforms such as Hadoop into existing data warehouses including data sharing and transfer into existing systems are some of the challenges. Security of data being shared for analysis and privacy of data for public consumption are also important. Country-specific rules for open data platforms and application specific sharing of data should be done. A skilled system for training for Data Scientists is still in its early stages in India. A major Challenge in implementation is to compare and make required data available to users in India which requires efficient techniques to store, manage, analyze and share such data. Policy regulations are needed to avoid the data wastage and preserve the privacy of the citizens.

A huge sum in the vicinity of \$17 billion was invested by the Indian government in the "Digital India" program, which was seen as a game changer for the nation. This together with the interest in Smart City Projects, where the Indian Government plans to create 100 smart urban communities and has already

Sanctioned \$1billion which is simply the seed money. CISCO also declared an investment of Usd1.7billion in the smart city initiatives in India. The key factor for big cities is the Big Data that can be salvaged to enhance resources and make operations of a city more effective. India and other countries like Japan, Qatar, India, China, Sri Lanka and South Korea, are planning smart city initiatives to oversee vitality and plan for further development. These initiatives will further accelerate improvement in keen social framework arrangements, the Internet of Things and machine-to- machine cooperation.

V. BIG DATA IN INDIA

There are many different ways by which Indian Government can use Big Data Analytics for implementing various policies and to bring changes in the existing system to improve it. To increase the participation of ICT's in governance and implementation of data analysis techniques, The National e-Governance Plan (NeGP) was formulated by the Department of Electronics and Information Technology (DEITY) and Department of Administrative Reforms and Public Grievances (DARPG) on 18 May 2006 by the Government of India comprising .Some of the projects of Indian Government which are biggest possible contributors of Big Data are as follows:

- V.1 In India, one of the biggest sources of big data can be **AADHAAR** scheme. Unique Identification Authority of India (UIDAI) is the authority of Government of India to provide unique identification numbers to each and every Indian. Over 304 million (30.4crores) number of AADHAARs had been issued as of 24-March 2013 and this figure is more than 25% of the overall population of India. This figure gives us a clear idea that how big the data would be at the end of a year, so processing and manipulating the data using big data analysis would be helpful for the government to plan the policies accordingly.
- V.2 **Digital India** is an initiative of Indian Government to integrate the government departments and the citizens of India. It aims at ensuring the government services are made available to citizens electronically by reducing paperwork.
- V.3 The **National Scholarship Scheme** was officially launched on 1 July 2015, it is one stop solution for scholarship process right from applying, verification, sanction and expenses to end beneficiary for all the scholarships provided by the Government of India. There is a common application for all scholarships, so the application process is simplified and the system will itself suggest the schemes for which the students can apply. Transparency will be increased as duplicate applications will be eliminated. Now, every scholarship has been made online, which will generate a large amount of big data. To manage and process this much amount of data, big data analytics will be considered as a boon.
- V.4 The **IRCTC** is one of the most frequent visited websites of the country. But, the process of procuring a ticket was proved to be one of the most annoying and complicated tasks ever because the server used to give up on many systems just a few seconds before 10 am, and then, comes back to life only when all the seats have been booked. Now, the question is how can we revive our IRCTC website so that it can function properly without server failure?

Big data can be one of the means to answer it. And, this can be achieved by using reliable big data tools for the system to serve more users logging on the site. So, the good news is that Indian Railways are working out on this for some time and had taken some major steps for overhauling the IRCTC by implementing a new technology which is Pivotal Gem Fire. Pivotal Gem Fire is a distributed in-memory database which is part of Pivotal Big Data Suite. The resulting system was tested and it has been found out that the time taken to book a ticket has come down significantly.

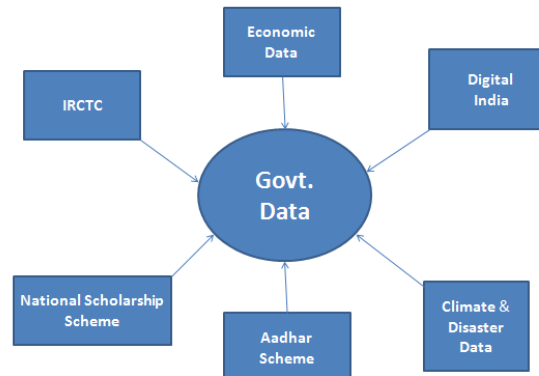


Figure 2. Sources of Big data for Government .

VI. CONCLUSION

In this paper, we studied that E-governance systems have large volume of data that cannot be processed and analyzed with traditional Relational Databases and if we analyzed this vast amount of data with good and efficient techniques and analytical tool, it can give rise to numerous ideas and effective results. Big Data Analytics plays a vital role in this crucial knowledge discovery from the mountainous amount of data available from smart cities and other resources and so is an indispensable requirement for productive e government. Although, the development and implementation of Big Data is difficult but researches are going in this area to make it useful to the country.

REFERENCES

- [1] A. G. Picciano, "The Evolution of Big Data and Learning Analytics in American Higher Education", Journal of Asynchronous Learning Networks, pp 9–20, 2012.
- [2] Ankita S. Tiwarkhede, Prof. VinitKakde," A Review Paper on Big Data Analytics", International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064.
- [3] Bansal, V. and J. Bhattacharya."E-governance solution for the government of Maharashtra. Technology whitepaper", India Research Lab, IBM, 2000.
- [4] Dean, J., & Ghemawat, S. "Map Reduce: Simplified Data Processing on Large Clusters". OSDI, 2004.
- [5] Hitesh H. Parmar. Application of BIG DATA in Government Sector. International Conference on Big Data Management & Cloud Computing, Bhubaneswar, ISBN: 978-93-83060-09-2, 28th July 2013
- [6] <http://hadoop.apache.org/>
- [7] Puneet Singh Duggal, Sanchita Paul, "Big Data Analysis: Challenges and Solutions", International Conference On Cloud, Big Data and Trust 2013, Nov 2013.
- [8] Shubham Kalbande, SumantDeshpande and Prof. Mohit Popat., " Review Paper on Use of Big Data in E-Governance of India", International Journal for Research in Emerging Science and Technology, vol2, MARCH-2015. E-ISSN: 2349-7610.
- [9] NeGP Website. "About NeGP". NeGP Website.
- [10] Gillick et al., 2006, "Map Reduce: Distributed Computing for Machine Learning", Berkley, December, 2006.