

# Big Data and the Opportunities and Challenges for Government Agencies

Varun Singh, Ishan Srivastava, Vishal Johri  
*Department of Computer Science and Engineering,  
 Manipal Institute of Technology,  
 Manipal, India*

**Abstract**—The data stored with Government agencies is an asset both for the nation and the government. This data which is a potential source of opportunity brings with itself many challenges and the government agencies like many other corporations should be able to seize the opportunity that this big data presents and utilize it to develop policies and deliver services to citizens. In this paper the authors have tried to highlight the opportunities presented to government bodies in relation to the use of big data and some other emerging tools and technologies which facilitate better appreciation of what this powerful yet unutilized information can tell us and also the potential threats it might pose.

**Index Terms**—Big Data, Big Data analytics

## INTRODUCTION

With the advancement in computer and communication technologies, the rate of the data generating and sharing has increased exponentially. Following is a list of some of the factors that have contributed to the growth of data production:

- *Increase in data processing capabilities*: Modern computers provide a significant increase in processing and storage capabilities. This enables the conversion of various types of content and media from conventional forms to digital formats.
- *Lower cost of digital storage*: Technological advances and the decrease in the cost of storage devices have provided low cost storage solutions. This cost benefit has increased the rate at which digital data is generated and stored.
- *Affordable and faster communication technologies*: The rate of sharing digital data is now much faster than traditional approaches. A handwritten letter might take a week to reach its destination whereas it typically takes only seconds for an e-mail message to reach its recipient.
- *Proliferation of applications and smart devices*: Smart-phones, tablets and newer digital devices, along with smart applications, have significantly contributed to generation of digital content.

Data can be classified as structured or unstructured based on how it is stored and managed. Structured data is organized in rows and columns in a rigidly defined format so that applications can retrieve and process it efficiently. Structured data is typically stored using a database management system (DBMS). Data is unstructured if its elements cannot be stored in rows and columns, which

makes it difficult to query and retrieve by applications. For example, customer contacts that is stored in various forms such as sticky notes, e-mail messages, business cards, or even digital format files, such as .doc, .txt, and .pdf. Due to its unstructured nature, it is difficult to retrieve this data using a traditional customer relationship management application. A vast majority of new data being created today is unstructured. The industry is challenged with new architectures, technologies, techniques, and skills to store, manage, analyze, and derive value from unstructured data from numerous sources.

The most widely accepted definition of big data comes from Gartner who defines it as high-volume, high-velocity and/or high-variety information assets that demand cost-effective, and innovative forms of information processing for enhanced insight, decision making, and process optimization [1]. These are better known as 3 “V”s of big data. Data analysts all over the world discuss big data in terms of its value – the economical/political value and its accuracy and the uncertainty. Governments store or have access to an enormous wealth of data which increases daily including, but not limited to, the spatial and location data, and data collected from and by its citizens. Research shows that this data, if utilized in a proper way, carries the potential to transform how the services are designed and delivered and transform them in such a way that they can meet the individual’s needs more accurately and in lesser time.

This improvement in the services can then diversify and reach out to areas like medical diagnostics, managing the infrastructure, personalized social security benefits, better emergency services reduced criminal activities spanning across government and non-government domains. New innovative services can be developed as the availability and usage of public sector information becomes prevalent.

Majority of private corporations hold and analyze data about their customers, and come up with new business models, services or products which eventually benefits their customers and which in turn is a benefit for them. Government Agencies should also cease this opportunity to learn from the innovative ways in which the private sector uses this data to operate more efficiently and bring more value to their customers. At the same time agencies should take full precautions to ensure that the issues relating to privacy and security have been addressed properly.

Companies like Google, Twitter and Facebook hold a lot of data of their users which are spread across the globe and

they offer this data on commercial terms. Government agencies can use this data by carefully considering its accuracy to begin with their analytics project. Of course, they should consider the ethical, privacy and security implications of such decision carefully before beginning any such project.

#### OPPORTUNITIES FOR GOVERNMENT AGENCIES

The government agencies have the opportunity to harness the potential of big data by employing adequate methods to unlock the value and gain an insight through transformation of the information, facts and relationships and indicators. Due to the limitations of the current technologies, this value which big data hides is limited because the agencies have limited ability to manage the 3Vs - velocity, volume and variety of big data and extract useful information. Also, each big opportunity is accompanied with some big challenges - and the agencies have to overcome them to fully realize the benefits of big data.

A careful examination of the advancement of big data technology has made it evident that it is a potential tool to enhance government agencies' analysis capability which, eventually, will enable the government to focus on services which are most valuable for their citizens. Research has made it evident that big data analysis can provide information of social networks and relations and provides the decision-making capability for a large number of applications.

Big data analysis is a buzz word in the recent times and every industry is trying to harness its capabilities. For government agencies, big data analysis can be useful to provide insight into key areas pertaining to society in general such as healthcare services, medicine, science and education, transportation and infrastructure services, communication, space research and meteorology, and also social sciences.

Some key areas in which the government agencies can harness the benefit of big data analysis are:

##### A. Management of data

By implementing practices which are streamlined with big data analysis, government agencies can save a lot of time and money which they spend on managing the data. Data sources from differing organizations and operational areas would be of greater benefit to multiple agencies and for multiple purposes if there were greater transparency. A simple example for this would be that the agencies would not have to spend time to redundant data collection process as they can reuse the data which was collected elsewhere.

##### B. Making services personalized

By using big data analytics, the agencies can recognize the services which bring value to an individual or a group. Big data can easily achieve this due to its inherent granularity - the ability to mix heterogeneous pieces of data chunks to generate useful information. This granularity may reveal a possibility of personalizing services which will be more beneficial for an individual and will be delivered by government.

##### C. Using predictive analytics in solving problems

By unifying the data from multiple sources combined with advance data analytics techniques will boost the problem

solving, which eventually, will lead to effective decision making due to improved ability of predictive analytics.

##### D. Increase in Productivity

By using big data analysis, the government agencies can identify cost savings. Opportunities to enhance the efficiency can be identified whose direct consequence will be an improved productivity. In totality, this will further boost innovation.

#### A SNEAK-PEEK INTO THE FUTURE

The government agencies should formulate strategies to harness the capabilities of big data. The strategies should be designed to highlight key opportunities and challenges that big data will bring to government agencies. The strategy should aim to assist agencies to take advantage of these opportunities and realize the potential benefits of these new technologies. Big data allows for more focused and evidence-based policy design and service implementation that in turn allows citizens to interact with the government in a personalized and seamless way. A successful big data strategy is expected to assist in realizing each of the priority areas which are:

- a. *Better services delivery* — the usage of big data analytics will allow agencies to provide personalized services that are designed to meet citizen's needs and preferences. For example, the government agencies can identify the individuals or groups who are eligible for certain entitlements without needing them to apply for it explicitly.
- b. *Efficient government operations* — by employing big data for predictive analysis government agencies will be able to assess risk and feasibility, and they will be more effective in detecting fraud and error. This will lead to enhancement in productivity as the government can engage more resources into the projects which have more impact and confidence of the outcome.
- c. *Increased Collaboration* — the usage of big data analytics and its related technologies will enable the government agencies to collaborate with industry, academia, non-government organizations and other interested parties locally and internationally. This will help in increasing knowledge, kick start ideas, spark innovation, and generate growth and formulate better decisions and solutions which meet the needs of local as well as international governments. Also, as the government agencies pursue big data technologies, it will open up a channel for collaboration between different agencies that will strengthen existing networks and help develop new partnerships.

#### CHALLENGES FOR GOVERNMENT AGENCIES

Along with the massive opportunities, big data brings with itself many challenges which the government agencies have to carefully address. The volume of data is already enormous and is increasing rapidly daily. Due to the proliferation of devices which have internet, the velocity of the data being generated is very high. In addition, the

variety of data which is being generated is diversifying and the agencies' capability to capture and process this data is limited.

Current software and hardware, technologies, architecture, management and analysis approaches are not enough to cope with the massive amount of data and agencies have to change the way they think, plan, work, govern, process, manage and report on data if they want to harness the full power of big data.

Some biggest challenges which the government agencies will face are:

#### *A. Challenges related to privacy and security*

The government and its agencies are committed to safeguard the privacy of its citizens. In doing so the agencies should set clear boundaries for the usage of personal information. Government agencies, when collecting or managing citizen's data must comply with the acts and regulations and other laws as applicable. The agencies should be careful to maintain public confidence in the government as a secure repository and chamberlain of citizen information.

The use of big data will introduce an additional layer of complexity in terms of management of information security risks. Big data sources, the transport and delivery systems within and across agencies, and the end points for this data will all become targets of interest for hackers, both local and international and will have to be protected. Many governments have open governance policies, under which they might release the large amount of machine readable data which could lead to a disaster as the data may provide important information to unfriendly state and non-state actors spread locally or internationally. This security threat will need to be understood very well and carefully managed.

The Big data gets its power by combining a number of apposite, different datasets which can be linked and analyzed as a whole to reveal new patterns, trends and insights. But before employing the big data analytics techniques, the government agencies should get the public trust and citizen's confidence. The government should convince its citizens that this linking and analyzing the data can happen without compromising the privacy rights. Government must ensure that the citizens' and public trust in government agencies is maintained. As the volume of data which government holds is increasing, the public trust can easily be affected by leakage of data or information into the public domain. Government and the concerned agencies have to consider this fact at the very beginning and should develop the secure systems with security as the prime requirement. Communication and collaboration with industry experts is an important first step to begin with.

#### *B. Challenges related to managing and sharing of data*

Accessible information is the lifeblood of a robust democracy and a productive economy [2]. Everyone who has worked on data analytics before will agree to the fact that for data to hold any value, it should fulfil 3 basic requirements - discoverable, accessible and usable. And the government agencies should realize that these

requirements become more significant when the focus of our discussion is towards big data.

Government agencies should work to achieve these requirements but they should comply with the privacy laws. The processes by which the government agencies collect, handle, utilize and manage the data should comply with all applicable legislative regulations and the government must focus on ensuring that its agencies analyze the data in lawful and meaningful manner.

In most of the cases, big data is utilized for complex analysis and to support decision making. For this to happen, the data sets which are considered should be accurate, complete and should be available on-time. Owing to these reasons, data movement and accessibility across the government agencies should be done via standardized APIs, formats and metadata. Following the standardized methods will produce better quality of data and it will produce phenomenal benefits as far as business intelligence is concerned.

Many governments these days follow open governance policy under which the data sets are available to public. The governments should also extend these 'open' initiatives to make data standardized, available and of course, open to flow within and between government agencies. This will result in increased collaboration among government agencies. But again, all this should be done to the extent permissible by the privacy laws. The inter-governmental agency collaboration brings the opportunity for agencies for innovation and competition in the marketplace. In order to promote the usage of government data across government agencies, the government can make the agencies to maintain information asset logs for information available to the public. These logs can be used by other agencies or within the same agency which will increase reusability of data.

In the process of making the data availability more fluid, the government and government agencies should consider that the new technologies can provide an opportunity for unfriendly state and non-state actors to gather or extract sensitive information from seemingly harmless data. There have been concerns that correlating the separate data sets that are individually unidentified, there are chances of extracting personal information. This is called 'mosaic effect' [3]. The agencies should take proper measures to ensure the anonymity and privacy is not compromised.

#### *C. Challenges related to technology*

The evolution of big data and its ability to tackle complex analysis of huge data, which seemed impossible in the past, can be credited to the recent advancements in the technology. If the government agencies want to embrace big data analytics, it will put a lot of stress on current hardware and software which is already burdened with tasks like processing, analyzing and storing data. The government agencies should manage these technological challenges and gaps efficiently in order to gain full benefits of big data.

Particularly, these technologies constitute low cost storage technologies and arrays, in-memory processing and cloud

based solutions clubbed together with new software products which are supported by high performance servers and processing platforms. The advancements in Cloud computing over the last few years is the prime factor of wide adoption of big data analytics. With features like tiered storage, compute and analytics capabilities, widely available software solutions and granularity in selecting services, cloud computing has made big data analytics approachable. The agencies should utilize the flexibility offered by cloud computing to store manage and perform computational analysis on ever expanding data in a manner which was not possible before.

The government agencies need to have network which will provide sufficient bandwidth to transfer the data and enable real time analysis of data in cloud environment. The feasibility of big data projects should be determined based upon the government agency's business needs rather than preconceived technological preferences. The Government agencies should collaborate with other vendors and developers of big data solutions, and work with them so that they can come up with more capable tools and technologies which can ease the challenges of big data analysis.

#### *D. Challenges related to skills*

Big data is relatively young, and fairly complex because of which government agencies have to hire employees with new and diverse skill sets. This includes people skilled in science, engineering, technological research, mathematics and statistics, analysis and interpretation, business sense and understanding of the underlying nature of the business process or policy intent and above all, creativity. It is unlikely that government agencies can find people having all the skills and hence agencies have to form teams of specialists to allow the government agencies to achieve the results which they desire from their data analysis efforts.

Industry experts have reported that there is a major shortage of data scientists and people having experience in big data analytics. As reported by Gartner, by 2015, big data demand will reach 4.4 million jobs globally, with two thirds of these positions remaining unfilled [4]. Currently, there is a shortage of degrees and courses whose curriculum is focused on big data analytics. The big data analytics industry is searching for qualified people who are skilled in big data analytics. Educational institutions and universities are pushed to come up with the courses that provide education and train people for this line of business. It is evident that it will take some more time before government agencies can find skilled people for big data analytics. This leaves the government agencies to seek support from the rich experience and expertise in big data available outside of government. Government agencies

should also strive for opportunities for better collaboration with academia and industry (scientists, vendors and solution providers) and also independent research institutions. This will allow the government agencies to retain skilled resources and attract more expertise.

#### **CONCLUSION**

Even though big data is still in its initial stages, it has shown that if utilized properly, it can bring a revolution. Government and the government agencies hold massive amount of data and by applying big data analytics on this data sets they can improve government operations, policy development and provide services which provide more value to the citizens. Before the agencies can work upon how to utilize the big data analytics benefits, they should ensure that the privacy issues are addressed up front and proper measures are taken to protect privacy rights of the citizens and all the privacy and security legislations are adhered to. In this paper we have highlighted few opportunities which the big data brings for government and its agencies. We have also highlighted some challenges which will come when the big data analytics will be used by the agencies.

#### **REFERENCES**

- [1] Gartner, Gartner Reveals Top Predictions for IT Organizations and Users for 2013 and Beyond, <http://www.gartner.com/it/page.jsp?id=2211115>
- [2] Office of the Australian Information Commissioner, Open public sector information: from principles to practice, [http://www.oaic.gov.au/publications/reports/Open\\_public\\_sector\\_information\\_from\\_principles\\_to\\_practice\\_February2013.pdf](http://www.oaic.gov.au/publications/reports/Open_public_sector_information_from_principles_to_practice_February2013.pdf)
- [3] Jaikumar Vijayan, Sidebar: The Mosaic Effect, [http://www.computerworld.com/s/article/91109/Sidebar\\_the\\_Mosaic\\_Effect](http://www.computerworld.com/s/article/91109/Sidebar_the_Mosaic_Effect)
- [4] Gartner, Gartner Reveals Top Predictions for IT Organisations and Users for 2013 and Beyond, <http://www.gartner.com/it/page.jsp?id=2211115>
- [5] Zikopoulos, Paul C., *Understanding Big Data*, McGraw-Hill, 2012.
- [6] Garside, W., Cox, B., *Big Data Storage for Dummies*, Wiley, 2013.
- [7] Smolan, R., Erwit, J., *The Human Face of Big Data*, Against all odds productions, 2012.
- [8] Gleick, J., *The Information: A History, a Theory, a Flood*, Knopf Doublebay publishing group, 2011.
- [9] Mayer-Schönberger, V., Cukier, K., Poston, R., *A Big Data: A revolution that will transform how we live, work and think* Houghton Mifflin Harcourt, 2013.