APPLICATION OF BIGDATA IN GOVERNANCE

Rishabh Mishra[1]

[1](Department of Computer Science,Assistant Professor,DSITM, Ghaziabad, UP, India,rishabhvgi@gmail.com)

Abstract—*we are talking about how big data technologies can change the way traditional and age old processes can be made much more simple, effective and cost efficient. Same way not only the giant companies deals with the massive amount of data but even the Government of any Country are canonical users of Big Data, as they also keep track of massive number of different records for their country which may include information about the people of the country, their growth and many more things. In this paper I am going to focus on some of the application of the Big Data in the field of Government, Public sector and the society. In general There are many ways in which big data can help the Indian Government to grow more and go for the changes and implementing the various policies and government schemes*

Keywords— BIGDATA, Application of Bigdata, MYGOV.IN, Data Scientist, BigData in Census.

1. INTRODUCTION

Gathering and processing of huge amount of data is not new, what is new is the speed at which we can now process that data. This is an era of collective intelligence. Every day we create approximately 2.5 quintillion bytes of data, it is estimated that 90% of the data which is been generated till now is created in last 2 years alone, and with the help of that thing we can predict what amount of data the world would be dealing within upcoming years, and that lead us to work with the term that is called “Big data”.

Big data is one of the Buzzing word all around now, there are lot of companies working on the processing and manipulation of these data, big data usually means the collection of data that is too big and massive generally in petabyte, and that data can come from many different sources like click streams, sensors, and many other places, it must be processed quickly and which is too hard to be managed by all the existing tools to process those massive amount of data records.

Webopedia defines “Big Data” is a buzzword, or catch-phrase, used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process using traditional database and software techniques. In most enterprise scenarios the data is too big or it moves too fast or it exceeds current processing capacity.

While the term may seem to reference the volume of data, that isn't always the case. The term big data, especially when used by vendors, may refer to the technology (which includes tools and processes) that an organization requires to handle the large amounts of data and storage facilities.

The term big data is believed to have originated with Web search companies who had to query very large distributed aggregations of loosely-structured data.

We understand how analytics can change our perception towards data. But here we are talking about how big data technologies can change the way traditional and age old processes can be made much more simple, effective and cost efficient.Same way not only the giant companies deals with the massive amount of data but even the Government of any Country are canonical users of Big Data, as they also keep track of massive number of different records for their country which may include information about the people of the country, their growth and many more things. In this paper I am going to focus on some of the application of the Big Data in the field of Government, Public sector and the society in general.

One of the biggest example of using big data in the field of Government and politics has been given by the president of United States of America, Barack Obama. President Obama used the information of the voters of the previous election, like their email addresses their phone numbers, and team also tried to get feeds of various activities of people through the social sites like twitter and facebook, and based on those feeds people used to receive emails regarding the presidential campaigns, what are the policies which are going to be developed which will help people in near future, what are the other facilities

people will get after the election, People were directly targeted based on their previous data and their activities and they used to receive emails which will help people to understand more about the campaigns, and that lead to victory of President Barack Obama in 2012, and reelected as president of USA.

Now we all know that elections are not solely by analytics. The policies of the two candidates contributed to the results, and there may be involvement of some other factors too.

The main thing is that how can we use big data thing within the government and use those information and data records do more and more with less and less expanses which will help country and people.

2. BIGDATA SCENARIO IN PRIVATE SECTOR

Across all four regions of the world that we surveyed, 53% of the 1,217 companies said they had undertaken at least one Big Data initiative in 2012. The U.S. was the leader among regions in Big Data use. Of the four regions surveyed, this region had the highest percentage of companies that reported at least one Big Data initiative in 2012: 68%. Only a third of the U.S. companies said they didn’t have even one Big Data initiative.(Fig-1).

Of the 53% of companies with Big Data initiatives in 2012 (643 in all), the Asia-Pacific region had the lowest percentage (particularly Australia and Japan) at 39%. A slight minority (45%) of European countries said they undertook Big Data initiatives and about half of the Latin American survey respondents were in this game. A country-

by-country analysis revealed a number of differences. A high percentage of Indian (70%) and Mexican (68%) companies reported Big Data initiatives. (Fig-2).

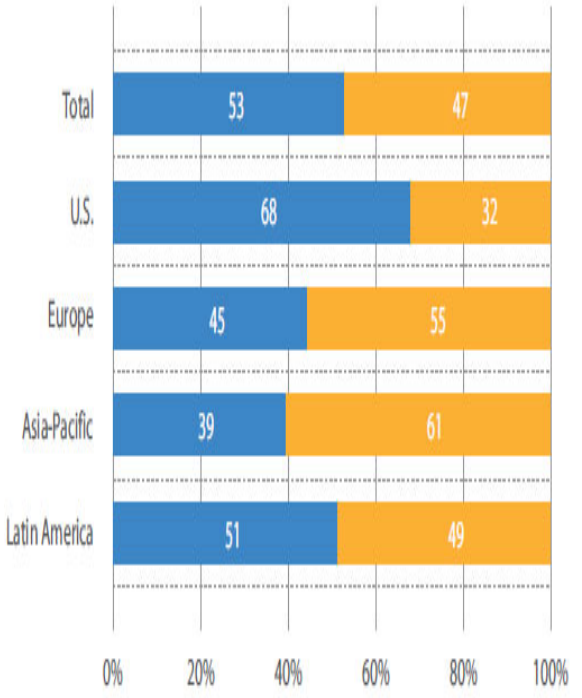

Fig-1: North American Companies are More Likely to have Big Data initiatives (Percentage of Companies with BigData Initiatives in 2012)

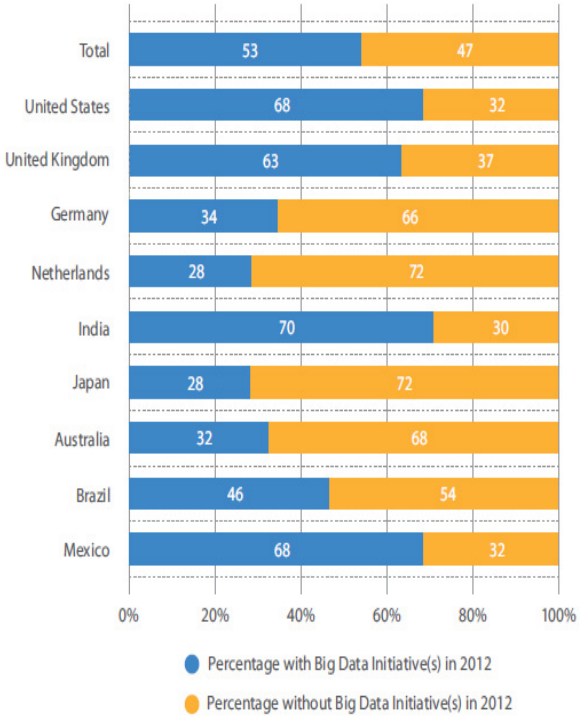

Fig-2: US, India, UK and Mexican Companies are More Likely to Have Big Data Initiatives. (Percentage of Companies by Country With Big Data Initiatives in 2012)

3. BIG DATA AND INDIA

There are many ways in which big data can help the Indian Government to grow more and go for the changes and implementing the various policies and government schemes. We understand how analytics can change our perception towards data. But here we are talking about how big data technologies can change the way traditional and age old processes can be made much more simple, effective and cost efficient. Let's take the case of Census in India, which is supposedly that largest exercise of its kind in the world.

Census in India works on de facto canvasser method. Under the 'Canvasser' method the enumerator approaches every household and records the answer on the schedules himself after ascertaining the particulars from the head of the household or other knowledgeable persons in the household. De-facto basis is counting the population at the place where a person is actually found on the reference date of the census rather than on the place of normal residence."

In the first phase of the census the enumerator visits the house hold with questionnaire for the survey. They move from house to house and collect the information manually. This practice is extended for 7 months of survey round and after that the information will go for revision round. In revision round any changes in birth, death and migration will be updated compared to the last time of visit. The survey itself has two phases' house listing and house census and population enumeration. In house listing each building related structures are updated in the survey and also other house amenities. In second phase of population enumeration every individual in the house hold will be updated through the survey. The next phase is the data processing phase. The processing of the census data occupies a very important place in the census. Data capturing and data processing is giant phase of census as census generate large amount of data. To convert data to electronic form, it scans census schedule and extract records from it with the help of high speed duplex scanners and reading information using ICR technology. This was the advanced approach which Indian census body adopted and it did paid them enormously as it reduced their time and cost. The data size from this process can be in tera bytes (TB).

After data processing, evaluation and analysis of results ensures that the completeness and accuracy of the published results. "The extent of error can be estimated through the use of checks of the internal consistency of the data, by examination of the reasonableness of the results and by comparison of the results with data collected in other enquiries. The publication of the census results must specify the extent of error in coverage and reporting. This will greatly help to make comparative studies and also indicate the long-term trends of certain characteristics of population. " The whole data processing is done in RDBMS (SQL Server – 2008) environment, where data is stored in relational tables and further queried for different summarizations.

Since this data is in TBs (tera bytes), web scale technologies like Hadoop and NoSQL database like Cassandra makes an ideal solution for this. Hadoop with its Map Reduce framework and very high scalability will make

the whole process complete in weeks and Cassandra can store this summarized data in query optimized tables which will make dissemination of results also faster.

New insights from this whole data can be generated in days. This will enable new research and analysis of data outside the census organization also faster and easier, enabling decision makers with vital data for proper planning. The whole 10 years exercise will come down to 2-3 years and the Census organization can do more frequent census as the rate of high rate of population and economic growth makes the data far from real in less than 5 years' time.

It looks Census in India is an ideal fit for a Big Data solution not only enabling quick decision making but also less costly utilizing the powerful open source alternatives.

Now if we consider India then one of the biggest source of big data can be of AADHAAR scheme, a unique identification project, which is undertaken by the Unique Identification Authority of India (UIDAI), is an agency of Government of India, which was established in February 2009, and will own and operate the Unique identification Number Database. The authority is planning to provide the unique identification numbers to all Indians, but won't be providing the smart cards, that would depend on to the ministries of the country. The authority will maintain a database of resident containing the biometric and data. Mainly the database would be containing the 12 digit-unique number which is GOING TO BE ISSUED BY Unique Identification Authority Of India for all the resident, all this number will be stored in the centralized database and will be linked to the basic demographics and biometric information-photo,10 fingure prints and iris of each individual .Random number generated will be free from any classification based on caste, religion and geography. This database will be very large and bid but this will allow those people to have identity who do not have sort of identification. Now if we talk about the number of entries under this scheme, total number of AADHAARs issued as of 24-March-2013 is over 304 million (30.4 crore). This is more than 25% of the population of India. And that gives us the insight that how big the data would be within the end of this year, and processing and manipulating this data statistics would help the Government to make various decisions based on those statistics.AADHAAR is an ambitious government Big data project which is going to be one of the largest biometric database in the world by 2014, with the global capturing about 600 million Indian Identities. And dealing with this kind of database is not an easy task, but India is going to be one of the first country who would be dealing with this much large amount of public biometric database.

4. BIG DATA AND ITS APPLICATIONS IN INDIA

A. Direct Benefit Transfer Scheme

IT is also known as the DBT scheme that aims bringing transparency and eliminating the stealing from the distribution funds. This scheme is dependent on AADHAAR card that will help people to get benefit of this scheme and it will also ensure the government about the number of people who are getting benefit out of this, it includes various different schemes under it like education, scholarships etc.

By keeping track of this records will help the Government to decide the funding for a particular states, and they would also be able to keep track of improvement and the growth within the particular region where people are benefited of this scheme.

So if these records are analyzed in a proper manner then this statistics will lead to the improvement, and Government would be able to focus on those areas which are really important, thus how big data can play a vital role there.

B. Impact on Election and Voting system

As I had mention the biggest example of Big Data and its usage in the Election campaign by the President of United States, Barak Obama, his team focused on to the right public or we can say registered users as per their database records and based on their activities user used to receive emails and connected through phone calls.

President Obama's election team was able to dynamically analyze the global state of pre-election sentiment, and deploy advertising resources andhuman volunteers to the areas that needed the most attention, virtually in real time. Indian Government can also use the same kind of thing, by analyzing the big data and making policies and the scheme based on those statistics which will help the people of the country as well as the growth of the country

C. Impact on Government Construction Projects

As there are many new projects are assigned in to the various states of the country every month for the development of the region and the country. It may be possible that Construction or the project may be deficient and may create problem in near future, as some of the projects might not be meeting the inspection criteria.

So my point here is that real-time analytics can help the government and the other people alike, let me give you an example for that, The University of Texas is working on sensor technology that use wireless sensors to identify failing bridges, lower the cost of monitoring those bridges, and improve the safety of new bridges, and can report dynamic data measurements on a bridge's condition. They're working on sensors that can survive the constant vibration, weather, and even send and receive data through all the steel that normally would make radio transmission a near impossibility.

So the thing would be like there would be lot data which is supposed to be filtered out to get the exact information about the bridges and their conditions, that will be bit complex but at a same time it will save lot of resources like money, human lives, etc. This data can be feed to the system and that system can help in the time of critical situation when the project need some maintenance or any repair kind of thing. Thus it can help in that way too.

D. Impact on Education

More data about students can help students in various different ways to find their interest, strength, and help to identify student weaknesses in ways that are not possible today.

Impact of Big data on to the Education sector can play a really vital role, in the normal case generally school boards

decide which text books students are required to follow for their particular subjects and all, but do we know which text book is actually good or bad? No, we depend upon expert opinions.

But now we do have the data, as are moving from the hard cover books to the electronic books, as people are using the tablets and the other electronic book delivery mechanism like ipad or Kindle from Amazon etc. They all usually collect the information. If we take an example of an Amazon then it know how long you stay on each page, when you re-read a page, or you read chapter over and over when you

Have some difficulties within that. All information is kept within Amazon. Those information would be really help a lot to the author or the publisher of the book to improve the things within text books, and that can also help the Board of Education to make some worthy decision based on those statistics, and that will ultimately help the students, same way to the people of the country.

E. Impact on Health care

This also a very important area where big data can play a very huge role and can save tons of money. By having an access to all the medical records of the past decade or more, can help government to take a proper decision in this area.

When this user data is processed in the right manner and analyzed correctly, it will result in to the complete statistics on which part of the country are suffering from what kind of health related issues and disease. And that can help the government to start new services and can also lead to opening the new hospitals to serve the patients, Government can also issue some more funding for the research and development of some lifesaving medicines, this way government can help in health care area.

F. Generating revenue from Government sites

With the Big data technology, Governments can also earn from their running web sites, as they are providing different kind of advertisement on the government sites, and many of the advertisement may be irrelevant.

For example site of Indian Railway Catering and Tourism Corporation limited (IRCTC), they deal with the tourism things, they may earn if they provide tourism relevant advertisement.

With the help of big data they can provide the personalized ads, which would be more relevant to the people who are using such sites, and that will usually increase the rate of generating revenue compare to those irrelevant advertisements.

G. Efficient banking system

Luckily, banks have a huge advantage — data. If pulled and analyzed, every transaction data can reveal tons of unmatched insights into customers' needs and deeds. Data represents a great competitive advantage and it can make banks smarter and more efficient. Today's tech savvy financial businesses and banks understand the value their data represents. They understand that when their structured operational data is integrated with the unstructured data from web, social streams can provide them with valuable insights on transaction patterns to predict customers' behavior in the future and to customize their business models to include ultra-personalized solutions suitable to their each and every customer. To do this, banks require support from high-tech big data firms that can help them combine their data with external data, cleanse and mine for valuable insights. The algorithms and technology required for doing this are highly sophisticated. Hence, it requires a different talent pool to work on, that very few banks can afford. There are many businesses who can cater to these specific needs of banks with their own proprietary platforms. A classic example is the big data driven B2B offering like Merchant Optimizer that helps banks not only to tap into a broader merchant network, but help them optimize their active network for profitability and effectiveness. Another example is Personal Choice app that helps to deliver ultra-personalized rewards & choices to the banking customer, based on their taste, influence, behaviour and context.

Even though customer analytics is the primary focus of banks, there are also vast options available to improve their analytics in areas like fraud detection, risk identification, mapping credit worthiness, regulations and compliance by using big data analytics

The banking sector is rife with change and unpredictability. Owing to its complex environment, it's difficult to predict the future of this sector with any degree of confidence how changes in banking laws and regulations will affect profitability, what should be the stress scenarios, what is needed to correctly measure each business line's different risk characteristics and where can we more effectively apply better customer models to reduce risk factors and financial losses. To find out answers to these puzzling questions and ambiguities, banks can use big data analytics to answer these and identify additional questions to effectively manage risk and drive risk-adjusted performance.Thus, leveraging big data effectively by integrating both internal and external data can turn routine transactional data intoa solid competitive advantage

5. SCOPE OF DATA SCIENTIST IN GOVERNMENT SECTOR

The demand for data scientists in India is growing in government agencies at par with private establishments due to a shortage of skilled manpower in the industry, an expert has said.

Data scientists, who analyse `big data` or high-volume, high-velocity data are required not only in industry sectors but are increasingly being drawn from the agriculture and clinical research areas as well.

Industry experts at the recently held "Big Data and Analytics Summit 2014" said India will be short of two lakh data scientists over the next few years.

In India, the "demand for data scientists in government agencies is on par with the demands in the private sector," There is a gap of skilled manpower in the industry. To address such gaps almost all large organizations are using such analysts. Such analysts are being used in a big way in government (for example, the National Population Register) and public sector as well as the data volume is huge.SAS or Statistical Analysis System is a business analytics and business intelligence software. The department of economics and statistics as well as the Maharashtra

government has been using such software and data scientists, he said.

To bridge the gap, SAS has tied up with one of eastern India`s leading business schools, the Calcutta Business School (CBS), to impart a certificate programme in business analytics for post graduate diploma (PGDM) students that would lead to a globally recognised certification.

6. INITIATIVE OF INDIAN GOVERNMENT IN BIGDATA

The Prime Minister's Office is using Big Data techniques to process ideas thrown up by citizens on its crowdsourcing platform mygov.in, place them in context of the popular mood as reflected in trends on social media, and generate actionable reports for ministries and departments to consider and implement. The Modi government has roped in global consulting firm PwC to assist in the data mining exercise, and now wants to elevate Mygov.in platform from a one-way flow of citizens' ideas to a dialogue where the government keeps them abreast of some of the actions that emerge from their brainstorming. Ministries are being asked to revert with an action taken report on these ideas and policy suggestions currently being generated on 19 different policy challenges such as expenditure reforms, job creation, energy conservation, skill development and government initiatives such as CleanIndia, Digital India and Clean Ganga.

With the PM inviting Indian communities in America and Australia to join the online platform, which he has termed a 'mass movement towards Surajya', the traffic handling capacity of mygov.in is being scaled up consistently.

PwC's global leader in government and public services Jan Sturesson said that the participative governance model being adopted through mygov.in could become a model for the developed world.

"The biggest issue for governments today is how to be relevant. If all citizens are treated with dignity and invited to collaborate, it can be easier for administrations to have a direct finger on the pulse of the nation rather than lose it in transmission through multiple layers of bureaucracy," he said, not ruling out the possibility of using the mygov.in for quick referendums on contemporary policy dilemmas in a couple of years.Within the 19 broad citizen engagement themes on mygov.in, there are multiple discussion groups focused on specific subsectors and themes. When it was launched in July, the site enabled brainstorming among its registered users around seven policy challenges. Users are allowed to sign up for four discussion groups in areas of interest apart from a group dealing with issues in their immediate vicinity.

7. CONCLUSION

Big data can be really helpful if the data or records are analyzed carefully and if we use that statistics in a right way, it can really help the developing country like India. Big data can help the Indian Government to grow more and go for the changes and implementing the various policies and government schemes.

REFERENCES

[1] From database to Big data, By : Ssm Madden – MIT http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6 188576

[2] What is Big data? http://bigdataspeak.wordpress.com

[3] How President Obama's campaign used big data to rally individual voters. http://www.technologyreview.com/featuredstory/508836/ho w-obama-used-big-data-to-rally-voters-part-1/

[4] http://articles.economictimes.indiatimes.com/ 2009-07-16/news/28448725_1_ration-cards-pan-cards-biometric

[5] "Unique Identification Authority of India" en.wikipedia.org/wiki/Unique_Identification_Authority_of_ India

[6] 5 ways big data will change lives in 2013 www.forbes.com/sites/sap/2013/01/09/5-ways-big-data-will-change-lives-in-2013/

[7] Direct Benefit Transfer http://en.wikipedia.org/wiki/Direct_Benefit_Transfer

[8] Big data helped OBAMA to win US elections http://rasiej.com/news/2013/1/15/big-data-helped-obama-win-the-election

[9] Sensors Monitoring Bridges http://www.ece.utexas.edu/aboutece/research_detail.cfm?id =17

[10] Big data and Education http://www.geekwire.com/2013/big-data-transform-politics-education/

[11] http://www.cognizant.com/InsightsWhitepapers/Big-Data-is-the-Future-of-Healthcare.pdf

[12] https://www.irctc.co.in/

[13] http://articles.economictimes.indiatimes.com
[14] http://sites.tcs.com/big-data-study/
[15] http://computer.financialexpress.com