



Big Data Analytics in Government and Organizations

Avni Gupta, Aviral Mishra, Dhananjay Gupta, Divyanshu Gupta

IT Department, ABES EC (UPTU)

U.P., India

Abstract— *This research paper discusses the dependency of government, public and private sectors on big data. We outline the changing role of governments and private industries and how they are attracting the customer, investors and vendors involvement. It includes the research surveys that various companies have done about Big Data and concludes with valid statistics. Organizations have started spending more in this area. It further shows various organizations, how they are growing with it.*

Keywords- *Jigsaw Academy, Bhartiya Janta party, Cloudera, Internet of Things, Wikibon*

I. 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[1].The digital economy has seen an exponential increase in the production of data, not least in government- collected data about citizens and businesses, organisations internal operations and its own interactions with external parties such as suppliers and communities. Some 81percent of top managers responding to a 2013 global survey by Bloomberg Business week Research Services (BBRS) strongly agree that big data is crucial to meeting their mission. 75% Real-time Big Data is helping government improve the quality of citizens lives. 25% Real-time Big Data is mostly for bureaucratic use. 3 in 4 IT Officials say Big Data is helping to improve the quality of citizens lives. 82% of public IT Officials say the effective use of real-time Big Data is the way of the future [2][3]. IBM says "By 2015, big data demand will reach 4.4 million jobs globally, but only one-third of those jobs will be filled." [4]

II. INDIA AS A BIG DATA SOURCE

According to State of the India CIO Survey, 40% of Indian IT leaders planned to implement big data analytics over the course of the year 2013- while 16 percent said they were already in the process of implementing it. NASSCOM expects the country's big data industry to grow from US\$200 million in 2012 to US\$1 billion in 2015. Gartner's "Big Data, Bigger Opportunities: Investing in Information and Analytics" survey found that 42 percent of IT professionals across all sectors reported investing in big data projects or planning to do so in the coming year, reported Campus Technology[5]. According to Jigsaw Academy[7], the industry will witness several instances of predictive analytics using Big Data systems. The Governments sensitivity towards Big Data and its potential will also increase in 2015 and more data will be available publicly. Gujarat will emerge as the fourth centre for Analytics after Mumbai, Bangalore and Delhi. The number of analytics companies based in Gujarat has been on the rise due to which there will be an increased demand in that region for trained data scientists[6].

III. BIG DATA IN POLITICS

Social media played a prominent role in the election. Contemporary campaigns use data in a number of creative ways, but the on temporary campaigns use data in a number of creative ways, but the primary purpose of political data has been and will be for the foreseeable future primary purpose of political data has been and will be for the foreseeable future providing a list of citizens to contact. Campaigns need accurate contact information providing a list of citizens to contact. Campaigns need accurate contact information on citizens, volunteers, and donors. Earlier this year, India's Bharatiya Janata Party (BJP) led a decisive and shocking victory in the Indian elections. Many have speculated that the partys pioneering use of big data and social media analytics played a decisive role in the BJP's and Narendra Modi's success[7,8]. Now that he's in power, it would appear Modi's innovative use of big data is far from over. Prime Minsters Office is using big data analytics to process citizens ideas and sentiments through the crowdsourcing platform mygov.in, as well as continuing to mine social media to get a broader picture of citizens thoughts and opinions on government action. IBM can help to achieve business objectives by applying big data solutions to many key government responsibilities including crime prediction and prevention[9].

IV. BIG DATA INITIATIVE

Department of Science & Technology (DST) has recently initiated a new programme to promote Big Data Science, Technology and Applications. For fostering research in this high potential emerging area, a BDI programme support

scheme is launched. Following are broad streams under which financial support shall be provided to quality proposals in Grant- In Aid mode.

- Support for R&D projects
- Support for Establishment of Center for Excellence in Big Data Analytics, Predictive technologies, Cyber Security etc.
- Support for national level Conferences/ workshops/ Seminars/ brain storming sessions etc.
- Support for Training programmes: Inhouse training programmes for Faculty/UG/PG/ Doctoral students

V. EMPOWERING THE DATA DRIVEN ORGANIZATION

Cloudera is revolutionizing data management with the first unified platform for big data, an enterprise data hub built on Apache Hadoop. Cloudera offers public sector entities a secure and cost-efficient place to store and analyse all their data, empowering them to derive new insights and correlation while extending the value of existing investments. The potential of big data for the public sector is unprecedented nearly 70% of all agency leaders believe that in the next five years, big data will play a critical role in fulfilling mission objectives.[18] Yet one third of that data is unstructured and more challenging to find and yield value. And often, these data sources are scattered and stovepiped. If government agencies could more effectively relate and analyse these petabytes of data, society would reap the benefits. But budgets are limited, and the technical requirements to analyse diverse and petabyte-scale data sets surpass the limits of conventional data management architectures[19].

A. NEW SOURCE OF INFORMATION- INTERNET OF THINGS

The network of intelligent devices which include sensors to measure the environment around them, actuators which physically act back into their environment such as opening a door, processors to handle and store the vast data generated, nodes to relay the information and coordinators to help manage sets of these components. McKinsey Global Institute reports a 300 % increase in connected machine-to-machine devices over past 5 years and a sheer 809% decline in microelectronics pricing which is anticipated to lead in 1 trillion more things connecting to the Internet across industries such as manufacturing, health care, and mining with a potential \$36 trillion cost saving in operating costs[20].

B. CLOUDERA

With Cloudera at the center of an agency's enterprise data hub (EDH), analysts and business users gain unprecedented visibility to and correlation of real-time and archival data. From building a full 360-degree view of a supply chain risk management or data exploration to powering advanced geospatial analytics and provisioning tactical, mission-critical capabilities, [18][19]an EDH with Cloudera has a range of computing frameworks for the mission, including interactive SQL, search, and machine learning. Cloudera's estimated revenues grew from \$56 million in Wikibon big data forecast from last year to \$73 million this year. It was announced in December of 2012, the impact of Cloudera's \$65 million boost was certainly felt in 2013. Cloudera is estimated to be operating at annual revenue run rates of more than \$100 million.[20] Today, they have more than 130 customers with names like AOL, CBS, eBay, and The Walt Disney Company, to name a few.

C. DEFEND AND PROTECT

Splunk software provides a big data analytics platform that operates across multiple organizational data silos to support security, IT operations and application management challenges.[21] Agencies use Splunk software to streamline traditionally resource-intensive tasks like troubleshooting, root-cause analysis, compliance and identifying security incidents.

D. HUMAN SERVICES- PUBLIC SECTOR

Big data techniques are also used by the Department of Human Services to improve service delivery, including creating more personalised services and detecting fraud and compliance issues. The new survey, qualitative interviews and report from Bloomberg Business week Research Services clearly shows that most public sector agencies are struggling to deliver better service to their citizens while contending with severe financial constraints. But a host of technology innovations can help ensure safety, further improve quality of life and increase confidence in government [22].

VI. BIG DATA SHAPING CUSTOMER BEHAVIOUR

Over \$28.1 billion rise is expected for the 'Internet-of-Things' market by 2020. The average media consumption on mobile is increasing dramatically, providing InMobi with access to more consumer moments with each passing day[25]. Access to data allows Inmobi to discover new consumer behaviors and identify new audiences to drive your marketing objectives with unprecedented accuracy. We do this at scale with more than 872 million uniques and 159 billion ad impressions every month on the InMobi network[26].

VII. BIG DATA- MARKET GROWTH

According to the big data number crunchers at Wikibon, spending on big data hardware, software, and services added up to \$18.6 billion in 2013. That's about \$500 million, or roughly 3 percent, bigger than it projected the market to be a year ago when it did its last forecast. On a year-over-year basis, actual big data spending grew 58 percent. [27]These are heady days to be in the big data market, to be sure. Where the wider, \$2-trillion IT industry gets by on less than 10

percent growth, the big data market is red hot, and plotting a growth chart that should make it a \$50-billion business by 2017. The smaller pure-play big data vendors are making inroads, megavendors like IBM, Hewlett-Packard, Dell, SAP, and Teradata still rule the big data roost. Big Blue remains at the top of the big data heap, with nearly \$1.4 billion in big data bucks. [28]The average organizations will spend about \$8 million on big data projects this year, according to IDG Enterprises Big Data report for 2014, which found that nearly half of organizations surveyed have already begun big data projects or have plans to begin one. The biggest enterprises are investing the most heavily in big data, but smaller firms are not far behind. According to the survey, the average organization today is measuring an average of 164 TB data. Over the next 12 to 18 months, that number is expected to balloon by 76 percent, to 289 TB. Nearly a third of the biggest enterprises, meanwhile, will be working with a 1 PB dataset by the end of 2014, IDG Enterprise found. According to IDG Enterprises survey, 27 percent of survey respondents plan to hire data scientists over the next 12 to 18 months, followed by data architects (24 percent), data analysts (also 24 percent), and data visualizers (23 percent). [29]Twenty-one percent are looking to hire business analysts and research analysts over that timespan, the survey found. It is not surprising that 70 percent of enterprise organizations are investing in big mid-sized business organizations. As part of its market-sizing efforts, Wikibon tracked and/or modeled the 2013 Big Data revenue of more than 70 vendors.

VIII. BIG DATA REDUCES COST

Based on the extensive usage data, Singapore's Land Transport Authority has been able to build more cost effective routes for cars, buses and trains, schedule the best timings and lowest fares. The remarkable aspect is that while doing so, it has also been able to reduce the cost of its revenue collection. 83% of Federal IT Officials say Big Data can save 10% (380billion) or more from the federal budget; or about 1,200 per American[11].

IX. BIG DATA AT ORGANIZATIONS

A. AMAZON

They were one of the early adopters and are the only company that have a patent that allows them to ship goods before an order has even been placed. The data points are wide-ranging and far more indicative of what a customer is likely to be genuinely interested in. Today recommendations are based on their wish list,[15] the items they have reviewed and what similar people have purchased this creates a very rounded profile of a customer and is a great example of predictive analytics being used to its full potential.

B. FACEBOOK

Facebook's unrelenting use of analytics has been one of the most prominent examples of Big Data potential to be overly intrusive. Their business model is built entirely around the extraction of our data[15]. By targeting very specific ads at the millions of people who use the social network everyday, it seems as if Facebook knows more about you than your friends and family, and worryingly, the truth is that it probably does.

C. OBAMA/ BIDEN POLITICAL CAMPAIGN

Once it had come to light that 64% of voters were using online forums in order to verify a claim that had been made by a candidate, they knew who they had to persuade and what they were unsure about. It also allowed insights to be leveraged in real time without much lag, in short it gave Obama an awareness of what people were discussing and a platform to speak to people about the things that were concerning them[15]. It gave him and his team a welcome push at a time when they needed it most.

D. NISSAN MOTOR CAMPAIGN

Nissan have a whole host of localised websites designed to help consumers determine which Nissan is ideal for them. By aggregating the data points from individual customers, Nissan were able to paint a vivid picture as to the vehicles which were in demand throughout a particular region – this means that advertising campaigns and production can be tailored to suit the needs of a region instead of just a country as a whole[15].

E. GERMAN WORLD CUP WIN

There is arguably no greater prize in sport than the FIFA World Cup. As well as grabbing the headlines for their slick football, Germany's use of data and analytics also caught the attention of many.[15] Using analytics, they cut down average possession time from 3.4 seconds to 1.1 seconds, a critical improvement that made all the difference when they defeated Brazil in semi-final and against Argentina, where Mario Gotze's goal in extra-time gave them the most prestigious prize in sport.

F. ACCENTURE

Big data is clearly delivering significant value to users, the vast majority (92percent)[15] of all users report they are satisfied with business outcomes, and 94percent feel their big data implementation meets their needs. Larger companies are more likely than others to regard big data as extremely important and central to their digital strategy.

X. CONCLUSION

Commoditization of High Performance Computing and mass storage in conjunction with cloud computing, open source software and platform interoperability made it possible to deploy data analytics techniques in order to cope with data

volume, velocity and variety and to provide the insight needed to really benefit from this data deluge. The value of data at our fingertips is largely underestimated and unexploited today and in almost every sector, including science, health, e-commerce, government, energy, environment, and manufacturing, many applications need to be developed in order to deliver the promise of Big Data. Our lives will consequently be changing rapidly and a whole new way of science and business will be added to existing ones. Correlations and predictions will pave their way into data analytics next to causation, modelling and theories. The biggest challenge does not seem to be the technology itself- as this is evolving much more rapidly than humans- but rather how to make sure we have enough skills to make effective use of the technology at our disposal and make sense out of the data collected. And before we get to that stage, we need to resolve many legal issues around intellectual property rights, data privacy and integrity, cyber security, exploitation liability and Big Data code of conduct. Like in many other technological areas, customs and ethics around Big Data possibilities and excesses take time to develop. Promises of Big Data include innovation, growth and long term sustainability. Threats include breach of privacy, property rights, data integrity or personal freedom. So provided Big Data is exploited in an open and transparent manner, delivery of the promise of Big Data is not far ahead of us.

ACKNOWLEDGEMENT

This work was supported by our mentor Associate Professor Mr. Ashwin Perti. We are grateful for his contribution and positive motivation which got us through. We would also like to thank the HOD and Assistant HOD for their support.

REFERENCES

- [1] http://www.informationweek.in/informationweek/interviews/296242/help-enhancing-indias-national-security?utm_source=reference_article
- [2] http://www.sap.com/bin/sapcom/hu_hu/downloadasset.2013-09-sep-23-16.closing-the-big-data-gap-in-public-sector-pdf.html
- [3] <http://www.mondaq.com/australia/x/317214/Constitutional+Administrative+Law/Big+data+and+the+public+sector+strategy+and+guidance>
- [4] <http://www.cutter.com/bia/fulltext/reports/2013/02/index.html>
- [5] <http://cacm.acm.org/magazines/2014/3/172509-big-data-applications-in-the-government-sector/abstract>
- [6] http://www.indiaonline.com/article/news-top-story/government-will-adopt-big-data-on-a-massive-scale-in-2015-jigsaw-academy-115010700148_1.html
- [7] https://www3.nd.edu/~dnickers/files/papers/nickerson_rogers.2014.pdf
- [8] <http://dataconomy.com/narendra-modi-first-prime-minister-use-big-data-analytics/>
- [9] <http://www-01.ibm.com/software/data/bigdata/industry-government.html>
- [10] <http://www.sap.com/customer-testimonials/public-sector/city-boston.html>
- [11] http://www.livemint.com/Opinion/LaTyG3Q111tuYa73XNMXsJ/The-promise-of-big-data-analytics.html?utm_source=copy
- [12] http://www.indiaonline.com/article/news-top-story/government-will-adopt-big-data-on-a-massive-scale-in-2015-jigsaw-academy-115010700148_1.html#sthash.h3GS6ojK.dpuf
- [13] http://www.datanami.com/2013/10/03/the_big_data_market_by_the_numbers/
- [14] http://www.livemint.com/Companies/34xxx6tyTfx00y7tMhUCrN/Indian-startups-sense-big-opportunity-in-big-data.html?utm_source=copy1
- [15] <https://www.linkedin.com/pulse/20140924071836-13965885-top-5-big-data-analytics-success-stories>
- [16] <http://www.enterprisemobilityexchange.com/content/enterprise-wide/big-data-analytics>
- [17] <http://insights.wired.com/profiles/blogs/the-power-of-big-data-and-mobility>
- [18] <http://www.cloudera.com/content/cloudera/en/solutions/industries/public-sector.html>
- [19] http://www.datanami.com/2014/02/12/big_data_market_growth_exceeds_expectations/
- [20] <http://www.sramanamitra.com/2013/03/11/clouderas-big-data-growth-story/>
- [21] <http://www.wsj.com/articles/big-data-software-firm-splunk-reports-rise-in-revenue-1424994178>
- [22] <http://www.theguardian.com/technology/2014/jun/13/big-data-how-predictive-analytics-is-taking-over-the-public-sector>
- [23] http://researcher.ibm.com/researcher/view_group.php?id=4977
- [24] <http://www.vccircle.com/news/technology/2012/08/09/india-innovation-fund-invests-personalised-analytics-startup-iken>
- [25] <http://indonesiabigdatacem.com/conference-agenda/>
- [26] <http://www.predictiveanalyticsworld.com/toronto/2014/agenda.php>
- [27] http://www.datanami.com/2014/02/12/big_data_market_growth_exceeds_expectations/
- [28] http://www.datanami.com/2014/01/16/big_data_spending_to_hit_8m_per_organization_idg_says/
- [29] http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2013-2017