INTRODUCTION

Janakiraman Moorthy

We don’t need more data weenies and we don’t need more strategic marketing planners. What we really need are more people with a foot in each camp who can make some sense of all of this new technology.

—Don E. Schlutz

We are living in an era of data deluge. Data that the human race has accumulated in the past one decade, far exceeds the data that was available to mankind during the preceding century. McKinsey & Co. foresees that the society is ‘on the cusp of a tremendous wave of innovation, productivity, and growth as well as new modes of competition and value capture—all driven by Big Data’. They also expect that different stakeholders such as consumers, companies and businesses are likely to exploit the potential of Big Data. Eric Siegel, founder of Predictive Analytics World, estimates that on an average day we accumulate 2.5 quintillion bytes of data. Another important character of the ‘datafication’, as Viktor Mayer–Schonborge and Kenneth Cukier call it, is that ‘Data can frequently be collected passively, without much effort or even awareness on the part of those being recorded. And because the cost of storage has fallen so much, it is easier to justify keeping data than discard it.’ As the cost of storage has fallen so much, it is easier to justify keeping data than discard it. As the cost of storage has fallen and computing power has increased, the size of data that was challenging before, can be easily handled with a desktop computer now. Several estimates about the accumulation of data have challenged our earlier imagination. Data scientists are increasingly using data quantities in Peta and Zeta bytes. Businesses, governments and developmental organizations—all are foreseeing that Big Data is likely to create value in multiple

1 Schlutz, Don E. (2012). Can Big Data do it all? Marketing News, 46(15), 9.
2 Manyika, J. et al. (2011). Big Data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, McKinsey & Co.
3 Siegel, E. (2013). Predictive analytics: The power to predict who will click, buy, lie, or die. Hoboken, New Jersey: John Wiley & Sons.
4 Mayer-Schönberger, V., & Cukier, K. (2014). Big Data: A revolution that will transform how we live, work, and think. London: John Murray.
ways. Such a huge potential and disparity in access to data can create considerable tensions among different sections of the society.

Let us look at the business perspective here. There is no doubt now that organizations, especially larger corporations have started accumulating large amounts of data. However, the data is unstructured, voluminous and of high speed. For example, the corporations have access to multiple sources such as call centre logs, client chats, SMS texts, Instagram pictures, Click Stream on the web, social media such as Facebook, Blogs, CCTV, RFID, Barcode Scanner, Geographic Information Systems (GIS), Genomics, Youtube, Internet of Things (IoT), and the list is further expanding. As Moldoveanu\(^5\) claims, the real problem with Big Data is not storage, analysis, etc., but, how do the organizations effectively and efficiently transform relevant data reliably into useful information? Perhaps, this is the culmination of a number of complex problems that the manager/client is more likely to face in the wake of Big Data deployment.

If we look at McKinsey’s Dominic Barton and David Court, or Thomas Davenport or Moldoveanu, the problem seems to be at two levels—one is individuals (data scientists) with relevant competency, capability and attributes and the second one is transformation of organization to harness the potential of Big Data. Hence, the organization may not be able to analyse and gain insights using traditional ways and means. Based on their research, Davenport and Kim claim that the organizations utilize Big Data in a profitable manner, distinguishing themselves from the traditional data analytical environment. The distinctions are: (a) focus of attention shifting from stock to flow; (b) data scientists, product development and process development gaining control over data analysts; and (c) the epicentre of data analytics shifting from IT department to core business functions such as marketing, operations and production.\(^6\)

Like other socio-technical phenomena, Big Data triggers both utopian and dystopian rhetoric. On one hand, Big Data is seen as a powerful tool to address various societal issues, offering the potential of new insights into areas as diverse as cancer research, terrorism and climate change. On the other, Big Data is seen as a troubling manifestation of Big Brother, enabling invasions of privacy, decreased civil freedoms, widening of inequality and increased state and corporate control. As with all socio-technical phenomena, the currents of hope and fear often obscure the more nuanced and subtle shifts that are underway.

**WHAT IS ‘BIG DATA’?**

The term Big Data is to a large extent vague and amorphous. As different stakeholders look at Big Data phenomenon from a multitude of perspectives, it is not easy to pen down a precise definition. Information technology professionals look at Big Data as large data sets that require supercomputers to collate, process and analyse to draw meaningful conclusions.

Francis Diebold was perhaps the first to use the term ‘Big Data’ in 2003 for the current phenomenon or explosive growth of data. He stated,

> Recently much good science, whether physical, biological, or social, has been forced to confront—and has often benefited from—the Big Data phenomenon. Big Data refers to the explosion in the quantity (and sometimes, quality) of available and potentially relevant data, largely the result of recent and unprecedented advancements in data recording and storage technology.\(^7\)

Douglas Laney\(^8\) of META Group (Gartner now) wrote a white paper entitled ‘3D Data Management: Controlling Data Volume, Velocity, and Variety’ which became the hallmark of attempting to characterize and visualize the changes that are likely to emerge in the future. Though for almost a decade, it was in oblivion, it gained popularity with Laney’s update, ‘The importance of ‘Big Data’: A Definition’. It has become Gartner’s updated definition—‘Big Data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision-making, insight discovery and process optimi-

---

\(^5\) Moldoveanu, M. C. (2013). The ingenuity imperative: What Big Data means for big business. *Rotman Magazine*, 59–63.

\(^6\) Davenport, Thomas H., & Kim, J. (2013). *Keeping up with the quants: Your guide to understanding and using analytics*. Boston, Massachusetts: Harvard Business School Publishing.

\(^7\) Diebold, F.X. (2012). On the origin(s) and development of the term ‘Big Data’. *PIER Working Paper 12-037*. Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

\(^8\) Laney, D. (2001). 3D data management: Controlling data volume, velocity, and variety. *Application Delivery Strategies*, META Group. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Da-

ta-Volume-Velocity-and-Variety.pdf
ization.' Though Laney did not explicitly use the term Big Data, by expanding the single focus of Diebold, he provided more augmented conceptualization by adding two additional dimensions. The Wikipedia definition of Big Data is ‘a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, analysis, and visualization.’

Given that the phenomenon is cutting across several domains, more characters are propounded and for the sake of simplicity, they are portrayed as extensions of ‘V’s. Veracity refers to data integrity and the ability for an analyst to trust data and confidently draw inferences and convert them into strategies. Value refers to economic or business value. As Kirk Borne puts it, such characterization with Vs is both fortunate and unfortunate. Borne has listed these Vs as challenges in deploying Big Data into any use. We use the compilation of Vs as the challenges in Big Data deployment.9

- **Volume:** Quantum of data generated, stored and used is explosive now. Several organizations provide different estimates and forecasts.
- **Variety:** For a marketing manager, data can now be generated through multiple channels. Apart from the conventional data sources such as market research, readership survey, television rating points, household panels, online click steam, etc., there are now the social media, such as, Facebook and Twitter, call centres, chats, voice data, video from CCTVs of retail outlets, IoT, RFID, GIS, smart phone, SMS, etc.
- **Velocity:** Real-time data is accessible in many cases such as mobile telephony, RFID, Barcode scan downs, Click stream, online transactions and blogs. The data generated from all such sources can be accumulated with the speed at which they are generated.
- **Veracity:** Authenticity of the data increases with automation of data capture. With multiple sources of data, it would be possible to triangulate the results for authenticity.
- **Validity:** The terms veracity and validity are often confusing. Perhaps the term validity should be understood as in the market research methodology that the data should be representing the concept that it is expected to represent.
- **Value:** Return on Investment and business value are being emphasized more than value for multiple stakeholders.
- **Variability:** Variance in the data is often treated as the information content in the data. With a large temporal and spatial data, there can be considerable difference in the data at different sub-set levels.
- **Venue:** Multiple data platforms, data bases, data warehouses, format heterogeneity, data generated for different purposes and public and private data sources.
- **Vocabulary:** New concepts, definitions, theories and technical terms are now emerging; they were not necessarily required in the earlier context, for example, MapReduce, Apache Hadoop, NoSQL and MetaData.
- **Vagueness:** It relates to the confusion about the meaning and overall developments around Big Data. Though it is not necessarily characteristic of the Big Data deployment, it reflects the current context. This may change and more clarity is likely to emerge in the future.10

Hence, Boyd and Crawford11 define Big Data as a cultural, technological and scholarly phenomenon that encompasses a closely intertwined amalgam of technology, analysis and mythology, wherein technology includes computing power, algorithms, analytical tools and techniques, databases, data warehouses; analysis refers to identification of patterns for making claims about the phenomenon being analysed; and mythology refers to the belief that bigger data sets can help in achieving higher form of intelligence and knowledge with the aura of truth, objectivity and accuracy.

TechAmerica Foundation defines Big Data as: ‘A phenomenon defined by the rapid acceleration in the expanding volume of high velocity, complex, and diverse types of data. Big Data is often defined along three dimensions—volume, velocity, and variety.’12

---

9 Borne, K. (2014). Top 10 Big Data challenges—A serious look at 10 Big Data V’s. Retrieved 15 January 2015, from https://www.mapr.com/blog/top-10-big-data-challenges-%E2%80%93-serious-look-10-big-data-v%E2%80%99s#VLk8Ly6mRYo

10 Quoted Venkat Krishnamurthy, Director of Product Management at YarcData.

11 Boyd, D., & Crawford, K. (2012). Critical questions for Big Data. Information, Communication and Society, 15(5), 662–679. Retrieved from http://www.tandfonline.com/doi/full/10.1080/1369118X.2012.678878 [pdf]

12 TechAmerica Foundation. Demystifying Big Data: A practical guide to transforming the business of government. Retrieved from http://www.techamerica.org/Docs/fileManager.cfm?f=techamerica-bigdatareport-final.pdf
The Foundation also qualifies that Big Data requires ‘advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information’. Andrea De Mauro, Marco Greco, and Michele Grimaldi\(^{13}\) reviewed a number of definitions and proposed a consensual definition—‘Big Data represents the Information assets characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value.’ We slightly modify by incorporating some new facets into the definition. ‘Big Data refers to information assets characterized by high Volume, Velocity, Variety, Variability with Veracity subjected to specific Technology and Analytical Methods for deriving Value with Virtue.’ The new character added in this definition is ‘Virtue’ to integrate ethical and piracy concerns into the deployment consciously as it is being increasingly recognized that these issues are an inherent part of the Big Data phenomenon.

**Big Data Technologies**

Big Data became a reality with the development of certain computing technologies such as Hadoop, HDFS, MapReduce, IoT, etc. Here we provide a simple explanation of these technologies, though it is not a comprehensive listing of technologies for a Data Scientist.

- **Hadoop** is a disruptive Big Data technology originally initiated by Yahoo to build an advanced search engine and process the generated data. Hadoop has evolved into a large-scale data processing environment. It has two components: MapReduce for data processing and Distributed File System (DFS).
- **HDFS** is Hadoop Distributed File System, which is a combination of massive parallel computing and tolerance to fault while running software on a commodity hardware.
- **Google’s Bigtables** is a very large distributed storage system for structured data. It can manage petabyte of data on more than thousands of servers. It is a ‘sparse, distributed, persistent, multidimensional sorted map. The map is indexed by a new row key and column key and a timestamp, and each value in the map is treated as uninterpreted array of bytes.’
- **HBase** started with an objective of strong large tables extending to billions of rows and millions of columns. It is an open source non-relational database designed based on Google’s Bigtables.
- **HIVE** is a data warehouse which facilitates querying and manages large volume data sets residing in the distributed storages. It is built on top of Hadoop to provide easy ETL tools, structuring different data formats, accessing data in Apache HDFS (HBase), and executing query through MapReduce. Hive uses a simple query language called QL similar to SQL (Structured Query Language).
- **CASSANDRA** is a scalable database.
- **PIG** is a high level data flow language. It can be used for expressing data analytical programmes along with capability for evaluating these programmes. It is capable of substantial parallelization for handling very high volumes of data. PIG is used for interfacing HDFS and ETL. PIG uses PigLatin a text based language.
- **ZOOKEEPER** is a centralized software coordination service with a single interface. It documents, tracks configuration information, naming, distributed synchronization and group services.
- **MAHOUT** is a scalable machine learning library. From April 2014, MAHOUT replaced MapReduce for algorithm implementation with its new code base. It is a richer and efficient execution than MapReduce.
- **YARN** (Yet Another Resource Negotiator) is a more distributed and faster Architecture.
- **METADATA** refers to ‘data about data’. Structural metadata contains data about database design, specification of data structure, containers of data. Descriptive metadata contains data about individual instances of data application and control.
- **NoSQL** (Not Only SQL) is a database environment which is non-relational distributed database system. It provides ease in speedy organization of data analysis with high volume of data, with desperate data types. Sometimes it is also referred to as Cloud database, non-relational database.
- **IoT** (Internet of Things) is interconnected uniquely identifiable embedded devices and software. Things can be a wide range of devices that are implants into the living organisms, or embedded in small and large machines such as cars, turbines, sensors in buoys, etc. Information generated from such interconnected devices can be used for large-scale efficient automation.

\(^{13}\) De Mauro, A., Greco, M., & Grimaldi, M. (2014). What is Big Data? A consensual definition and a review of key research topics, presented at the 4th International Conference on Integrated Information.
Recommendations for Harnessing the Benefits out of Big Data

To capitalize on customer relationship management (CRM) and analytical technologies, organizations need to acquire competence and change the system and the processes. Dominic Barton, Global Managing Director and David Court, Head of Analytics Practice of McKinsey, propose that the organizations need to develop three mutually supportive capabilities:

- Source both internal and external data creatively by organizing IT infrastructure and identifying, combining and managing the data optimally.
- Mobilize competent manpower in different departments who can work with advanced analytical methods for forecasting and predicting, creatively generating and delivering value offers, and optimizing the overall processes dynamically.
- Transform the organization to harness the benefits of emerging opportunities due to Big Data.\(^\text{14}\)

Davenport and Kim suggest three logical steps to take advantage of Big Data. They describe the steps—framing the problem, solving the problem and communicating and acting on results—in their book in detail.\(^\text{15}\)

In framing the problems, the company should consider all stakeholders’ perspectives. It is important to develop consensus and buy in for the processes and outcomes of the analysis. Solving the problem includes decisions about choices of data, analytical methods, models, algorithms, etc.; the major shift is from analysing structured data analytics to combining structured and unstructured data such as text, voice and images. The third step is communicating and acting on the results; it is also equally important to implement the results both inside the company and in the market.

Jake Sorofman and Andrew Frank of Gartner, warn that ‘data alone isn’t what makes marketing more the needle for business’.\(^\text{16}\) They argued that for successful marketing both intellectual and emotional part of managers need to be combined judiciously. Their ‘Intelligent Brand Framework’ combines both data-driven and human-centric approaches for making marketing investments in creative and operational levels (Figure 1).

![Figure 1: The Intelligent Brand Framework](https://hbr.org/2014/02/what-data-obsessed-marketers-dont-understand/)

**Figure 1: The Intelligent Brand Framework**

| Centricity | Data | Human |
|------------|------|-------|
| Level of Decision Making | Strategic | Observation | Inspiration |
| Operational | Automation | Engagement |

Source: Adapted from HBR.Org.

Human Centricity is based on decisions which are merely based on the intuition of managers. They may use data for basing their intuitions. The decisions are outcomes of a combination of human emotions, intellect, inspiration and subjective judgement. On the contrary, Data Centricity refers to the process of decision-making wherein managers use data for pattern identification, predictions and assessment of achievement of outcomes. Strategic decisions are generally of long term in nature and operational decisions are related to the question of ‘how’ to deliver the value offers and experience to the customers.

Inspiration is capturing ideas, thoughts and insights which are indexed and stored for converting them into strategic advantage. Observation is the means such as focus groups, surveys, ethnography, text analysis, etc., that can be used for gaining insights into consumer behaviour. Automation of decisions is achievable by machine learning algorithms and data analytical tools. Automation can achieve speed, accuracy and ease of assessment of efficiencies for targeting offers to consumers. Engagement refers to humanizing brands by continuous communication with customers through various channels of communication such as advertisement, social media and placements. They recommend that the organizations can assess where they stand in these quadrants and the optimal position is maintaining a balance across all the four quadrants.

Let us look at an experience narrated by Christine Armstrong of Jericho Chambers.\(^\text{17}\) She was flying from Portugal to London in Business class on a BA flight. The flight steward communicated to her that they were short

---

\(^{14}\) Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, 90(10), 78–83.

\(^{15}\) Davenport, T. H., & Kim, J. (2013). *Keeping up with the quants: Your guide to understanding and using analytics*. Boston, Massachusetts: Harvard Business School Publishing.

\(^{16}\) Sorofman, J., & Frank, A. (2014). What data-obsessed marketers don’t understand. Retrieved 31 January 2015, from https://hbr.org/2014/02/what-data-obsessed-marketers-dont-understand/

\(^{17}\) Armstrong, C. (2015). Why I worry about data. Retrieved 30 January 2015, from http://www.jerichochambers.com/polemic-why-i-worry-about-data
of meals and as she had the fewest loyalty points on the flight, they can’t serve her lunch. Though it is a clear data driven decision, it is devoid of reality. Perhaps the flight attendants are not empowered to take in to the reality that, ‘she was six months pregnant’; or are they just an integral component of a computer/data analytical engine? One of her fellow passengers resolved the issue by donating his meal to her.

John Bradshaw, the author of Reclaiming Virtue, claims that practical wisdom ‘is the ability to do the right thing, at the right time, for the right reason’. It is based on Aristotle’s belief that ‘practical wisdom was the virtue that made all the other virtues possible. Without the correct application of practical wisdom, the other virtues would be lived too much or too little and turn into vices.’

Waze, a mobile application, integrates GPS, social media networks and active or passive participation of the individual driver reports to create a visual advise/guidance application. It wants to improve the quality of driving by providing real-time information on the map. By connecting drivers to one another, we help people create local driving communities that work together to improve the quality of everyone’s daily driving. That might mean helping them avoid the frustration of sitting in traffic, cluing them in to a police trap or shaving five minutes off of their regular commute by showing them new routes they never even knew about.

By connecting drivers to one another, we help people create local communities of drivers that work together to improve the quality of everyone’s daily driving. That might mean helping them avoid the frustration of sitting in traffic, cluing them in to a police trap or shaving five minutes off of their regular commute by showing them new routes they never even knew about.

So, how does it work?

After typing in their destination address, users just drive with the app open on their phone to passively contribute traffic and other road data, but they can also take a more active role by sharing road reports on accidents, police traps, or any other hazards along the way, helping to give other users in the area a ‘heads-up’ about what’s to come.

In addition to the local communities of drivers using the app, Waze is also home to an active community of online map editors who ensure that the data in their areas is as up-to-date as possible.

Introduction to Colloquium Contributions

Rangin Lahiri and Neelanjani Biswas provide an account of Big Data application in the marketing domain. Marketers globally have always been challenged as to how to reach the right audience, how to get a prospect of who has a higher chance of becoming a customer or how to make a customer come for repeat purchases. As we move forward to the new dimension of data-driven marketing strategies, more and more complex business decisions are being made based on data, rather than on previous experiences and marketing hunches. This document elaborates how Data and particularly Big Data concepts can be utilized to make sound marketing decisions. The document also provides sound industry-specific business cases that can substantiate the use of data for marketing strategies with top and bottom line growth for an organization.

Dipyaman Sanyal and Jayanthi Ranjan explore the potential of Big Data applications in the government. Historically, governments have been some of the largest repositories of data and have utilized traditional data analysis techniques to dictate public policy and aid governance. However, with data volumes reaching Exabytes, the scope of data analysis has expanded exponentially. While some countries, like Singapore and the United States, have already started the use of Big Data analytics to help aid their decision-making, India has barely begun the process. As the Big Data initiative recently launched by the Government of India is still in its early stages, this article considers potential policy uses for Big Data Analytics by governments (with a focus on India) while also scrutinizing comparative policy uses by different countries and their relative successes and failures. Moreover, since governments have easy access to confidential data, questions of data ethics are examined.

Krishnadas Nanath integrates emerging new technologies—Cloud Computing, Big Data and marketing possibilities. With the growth of technologies like Cloud Computing and Big Data, many people believe that their implementation has become mainstream and

---

18 Bradshaw, J. (2009). Reclaiming Virtue: How we can develop the moral intelligence to do the right thing at the right time for the right reason. New York: Bantam.

19 Retrieved 27 January 2015 from https://www.waze.com/about

20 Moldoveanu, M. C. (2013). The ingenuity imperative: What Big Data means for big business. Rotman Magazine, 59–63.
that they are no more buzz words. However, it is important to hold on and make a reality check of how these technologies are marketed and implemented in today’s world. While many organizations do implement these technologies, several firms believe in mere association of names like Cloud with their products and services. Reasons for this could include lack of standards or just the pressure of implementing Cloud and Big Data in the industry. This article brings out the prospect of positioning Cloud/Big Data with a reality check on status quo and then leads to possible solutions to avoid the scenario of Greenwashing.

Pulak Ghosh and Janakiraman Moorthy discuss the customer privacy issues with the emergence of Big Data phenomenon. Given the nature of Big Data, information collected are likely to be in the memory for a much longer time than what it used to be earlier; novel and innovative means for combining and ‘sense making’ are likely to be deployed more and more; and a variety of structured and unstructured data would be combined. All such dramatic changes are likely to provide more powerful ammunition in the hands of business analysts and data scientists, and customers are likely to experience more invasion. New ways of handling such emerging privacy issues need to be developed.

A Perspective of Big Data in Marketing

Rangin Lahiri and Neelanjan Biswas

IMPORTANCE OF BIG DATA IN AN ENTERPRISE PLATFORM

Determining the real value of the information or data collected for decision-making is one of the biggest questions marketing intelligence groups are trying to identify. Big Data is one of the significant inputs for data; however, owing to the disparate channels and sources, it is imperative to differentiate between what information is useful and what is not for the purpose.

Figure 2: Sources of Big Data

Concepts of social media data, personalization of customer interfaces, better management and storage of structured and unstructured data are emerging strong. Social media data from Facebook, LinkedIn, Twitter, Google circles are being used proactively to understand the broader customer mood, swings and preference to particular brands.

The biggest complexity of managing Big Data is the fact that it is generated from many sources and in different formats as shown in Figure 2. Storing the data and performing analytics on the same makes it a more complex activity.

Figure 3 depicts how organizations are using Big Data for the purpose of understanding customers better and providing better products and services.

A single Big Data platform is now enabling Marketing Intelligence and Business Intelligence divisions churn out meaningful analytics which can drive business decisions. Big Data capabilities are helping large data set crunching for insights which were never understood earlier due to lack of measurement techniques and data collection avenues. Big Data is thus enabling business transformations purely based on data-driven conclusions.

Figure 3: Use of Big Data by Organizations
WHAT IF I KNEW?

‘If I knew’ is a particularly effective approach to discover opportunities in Big Data. It is the responsibility of the Chief Data Scientist to educate the rest of the organization and more importantly the marketing group on how to utilize data to boost revenue. Big Data is primarily used to solve specific business problems; however, knowing what data is available can help us think out of the box and use it creatively. This can help marketing groups to understand and unravel new opportunities.

‘What if I knew’ approach is helping organizations to:

- Analyse data and find out new patterns that yield new product segments and features. At times new products are introduced based on these data patterns. These result in new marketing opportunities. Analyses like ‘People who bought also bought.’ are enabling, among other things, bundling of products, discount campaigns and test marketing for new products.
- Retain and Xsell/Upsell existing customers—in fact it has been one of the foremost business cases of Big Data in Marketing. Enabling customers to put in review comments, rating products/services, discussing in social sites and empowering customers can act as effective strategies for retention and upsell opportunities.
- Empower a move to one-to-one marketing based upon consumers’ actual behaviours and preferences; additionally, it can also help to find new audiences and new groups/segments with different interests. Behavioural mapping can identify new customer clusters moving away from traditional customer groups. This can result in identifying new customers’ segments.
- Understand better how to target precisely a particular group of customers/prospects and which media to use for communication. This drastically reduces the expenditure for marketing and helps to get better responses from the target segment.
- Measure marketing channel effectiveness and campaign results. By funneling data on impressions, clicks, conversions, social actions and more into attribution and media-mix models, results can be obtained to measure effectiveness. It has already been proved that data-driven decision-making helps high-performing businesses.

A CASE IN POINT

Case Study: How a leading French Telecom Operator embraced Big Data to reduce operational cost, increase operational feasibility and enhance cross-sell/upsell opportunities.

Business Problem: The telecom operator was facing high operational customer service cost and falling cross-sell/upsell revenues. The CRM system had 10,000 users from 15 different source systems servicing 38 million service requests per year.

The Solution: Flexible architecture based Enterprise Data Management solution with Master Data Management and Big Data applications. This was clubbed with proven Search and Analytics platform with most connectors available off the shelf.

SOLUTION BENEFITS

- Operational efficiency for frontline customer service agents and marketing group
- 360 degree customer view enabled by Master Data Management
- Better cross-sell and upsell based on customer profile information availability and analytical predictions
- Access to all relevant information by one single application
- Reduction in churn by anticipation and qualifying customer expectation
- Increased customer experience resulting in reduction in average call length (ACL), reduction in call back rate (CBR) and increase in first call resolution (FCR)
- Consistent service between all channels and accuracy of information for all customer interactions
- Lower IT cost due to centralization of data, better data quality and analytics results based on single source of data
- Flexibility to upgrade, increase source systems, incorporation of new business rules, new products and changes in analytical rules.

FRONTLINE BUSINESS BENEFITS

- Customer Service Agents reduced average customer interaction time by 30–40 per cent as they had all information available on their screen and did not have to ask questions to the customers for information. This resulted in every agent handling 30 per cent more customers per day.
• Analytics provided information regarding customer household, product holding, rate plans, usage pattern, etc. Agents were enabled to actually cross-sell and upsell products to customers over the same service call. Over six months, this solution resulted in significant sales from a customer service office.

• Customer’s awareness and satisfaction went high as the service desk not only had all the information required but ready solutions over the call. One Bill services tied customers to single billing for Phone services, Data services and Digital TV subscriptions.

• Big Data crunched data across social media resulting in speedy resolution of all problems, sudden failure of services in specific circles, rewarding correct feedback providers and more.

• Analytics identified segments of customers who contribute the highest percentage of profits as well as minimum profit contributors. Better customer experience management initiatives were rolled out to reward better customers and increase loyalty management initiatives.

• Predictive analytics initiatives helped manage Risk and Control with better forecasting of revenue expectations.

• Return on Investment is expected to be 60 million euros over the next five years.

CONCLUSION

Big Data and Analytics are changing the way businesses are taking decisions. From traditional Business Intelligence and Analytics, the world is changing more towards Predictive Analytics and Big Data centric business decision models. But the biggest questions are:

• How business can benefit from Big Data solutions?
• Can Big Data remove focus from the core business problems?
• What would be the kinds of insights and predictive models that Big Data can produce?

Answers to these questions can differ from industry to industry as well as from one organization to the other. Some of the steps to understand the impact of Big Data can be summarized in the following activities:

Business assessment: understanding the business needs and objectives to define the Big Data POC use case/hypothesis.

Data assessment: auditing the real world information option available, both internally and externally, to define what sets of information are relevant to address the POC challenge.

Solution development: loading the identified data sets and analysing the information in respect to the case/hypothesis defined for the POC.

Insight: creating visualizations which demonstrate the value of Big Data POC on real business data using analytical models.

Instead of taking a big bang approach, Big Data has shown results with incremental approach or taking baby steps. The best way perhaps is to break it down into smaller pieces as shown above and then work on an enterprise level implementation.

The business assessment or proof of concept (POC) is more in the lines of consulting engagement that helps to explore opportunities provided by Big Data with minimal upfront investment. It is more like a modular engagement that lasts generally 6-8 weeks based on the scale, complexity and goals of the POC. POC helps to test particular business cases/hypothesis within an agreed scope. An indicative approach is demonstrated in Figure 4.

Figure 4: An Indicative Approach to POC

The end result is more like a full report of the findings from the POC along with simulation results.

The biggest advantages of POC or Business Assessment can be given as follows:

• Involves low commitment and low engagement, and is highly rewarding
• Has minimum cost impact and Big Data use cases can be simulated for understanding if that is really the approach a particular organization is looking for
Big Data Analytics for Government

Dipyaman Sanyal and Jayanthi Ranjan

To begin a discussion on Big Data Analytics (BDA) for governments, we need to first define BDA and then distinguish it from Big Data (BD) itself. We use a simple definition of Big Data in this article—as a generic term for data whose size and complexity is such that it cannot be analysed using traditional relational database methods. This implies that Big Data technology needs to allow for parallel processing of information in multiple servers and thus, the definition of Big Data is ever-changing since according to Moore’s Law, the computing power of devices will double approximately every two years.

However, this definition also indicates that volume alone is not adequate to define Big Data; the analysis of the data is a crucial component. Thus, technology needs to address not simply the storage of large data sets but also its exploration. Experiments on database management have highlighted the difference between the computing powers needed for storage and that of analysis. Jacobs used PostgreSQL to compare computing times for data storage and data queries. His experiment demonstrated that as data volume increases, the time required to perform a query increases by a factor greater than one. Since more than 80 per cent of the data is estimated to be unstructured, which is expected to double every three months, and as companies and governments move to analyse unstructured data from non-traditional sources that require more than simple queries and data slicing, this non-linear rise in computing time becomes particularly relevant. Despite substantial analysis costs, governments (or corporations) cannot ignore this data deluge any longer.

Moreover, BDA is evolving in tandem with the IoT. IoT makes it possible for ‘things’ (not including computers, phones or tablets) to be connected to other ‘things’, computers or mobile devices, which provides a vast array of information about choices and decisions. Currently, devices interlinked via the IoT number 12.5 billion and by 2020, this global network is estimated to reach 26 billion. In comparison, the number of computers and mobile devices in 2020 will be 7.3 billion.

Following the lead of larger corporations, national governments have started investing in Big Data and there is a growing awareness within the public sector that Big Data can provide significant support to policy making. According to the Tech America Foundation and SAP Survey (2013), ‘82% of public IT officials say the effective use of real-time Big Data is the way of the future’.

Globally, the utilization of Big Data technology and analytics for government is not a theoretical debate any longer but is now in the early stages of a practical implementation. In this article, we hope to address India’s tentative foray into the global Big Data community. Looking beyond the realms of technology, we address issues of decision-making, policy implementation, measurements, monitoring and ethics.
BIG DATA AND THE GOVERNMENT OF INDIA

Data can enable government to do existing things more cheaply, do existing things better and do new things we don’t currently do.25

In his seminal paper, 3D Data Management, Doug Laney26 delineated Big Data by identifying the three Vs that make up its defining characteristics—Volume, Velocity and Variety. Since these three Vs can unquestionably characterize the massive data that the Government of India needs to analyse, it is crucial for the government to exploit Big Data technology and analytics to help with good governance and shape its public policy.

- **Volume:** According to the 2011 Census, 1.21 billion people (about 17.5 per cent of the global population) reside in India. Keeping in mind the sheer size of the population, the task of creating a Unique Identification (UID) system to identify individual citizens and then collate the billions of data points of this population, is indeed a mammoth undertaking. To put this in perspective, the entire US census data from 1980 (population 226.5 million) occupied 100 GB of memory space.27 With six times that population, the size of the census data alone ensures that the tenets of Big Data technology are essential to the government. To gain insight from Big Data Analytics, the government will then need to interlink that vast dataset with myriad other data sources.

- **Velocity:** India has the second largest number of registered Facebook users (101.8 million) in the world after the US. It is also estimated to have the third largest Twitter user base (18.1 million) at the end of 2014.28 Together with data from traffic signals, cell phone towers, weather satellites, CCTV cameras and millions of other data generators, the velocity of data that the Government of India will need to deal with, especially in times of emergency, are enormous.

THE CORPORATE–GOVERNMENT DIVERGENCE ON BIG DATA USE

A proliferation of inexpensive data analysis tools and the possibility of information gathering at a fraction of the costs of household surveys have encouraged corporations to advance towards data-driven decisions. Although governments have lagged behind, implications and conclusions from data analysis by the government will have far reaching consequences due to its ability to dictate public policy. As Claire Vyvyan, Executive Director and General Manager of public sector at Dell UK, wrote in the *Guardian* last year, ‘Joining up public sector data sources can make government more efficient, save money, identify fraud and help public bodies better serve their citizens.’

Governments have traditionally collected data on its citizens to help it in governance and in policy formulation. From population figures to household income levels to commodity prices, data-driven decision-making has been the backbone of modern government economic policy. It is imperative that governments now adopt new technologies like Hadoop or NoSQL and develop advanced decision-making frameworks to transcend traditional data storage and analysis. Additionally, governments not only need to focus on collating data from the various sources it has access to, but also initiate organizing and analysing unstructured data received from physical files, scattered databases, legacy systems, tweets, videos, mobile phone signals, emails and GPS coordinates.

INTERNATIONAL DEVELOPMENTS

Despite the significant initial outlays necessary to set up a Big Data framework and the costs associated with data collation, governments across the globe are moving towards BDA due to longer term cost savings and prediction efficacy. Since President Barack Obama announced the Big Data Research and Development Initiative in

---

25 Tom Health Open Data Institute, quoted in Rutter, T. (2014, April 17). How Big Data is transforming public services—Expert views. *Guardian.* Retrieved from http://www.theguardian.com/public-leaders-network/2014/apr/17/big-data-government-public-services-expert-views
26 Laney, D. (2001, February 6). 3D data management: Controlling data volume, velocity and variety. *Application Delivery Strategies, Meta Group.* Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf
27 Jacobs, A. (2009). The pathologies of Big Data. *Databases ACM Queue, 7*(6).
28 *eMarketer.* (2014, May 27). Emerging markets drive Twitter user growth worldwide. *eMarketer.* Retrieved 30 November 2014, from http://www.emarketer.com/Article/Emerging-Markets-Drive-Twitter-User-Growth-Worldwide/1018674#sthash.pThElWgk.dpuf
2012, the US government has been one of the leaders in government expenditure on Big Data. To launch this initiative, six departments and agencies of the US Federal Government announced additional budgets of more than $200 million to improve Big Data tool and techniques. The plan for the initiative had three broad dimensions:

- Advance state-of-the-art core technologies needed to collect, store, preserve, manage, analyse and share huge quantities of data.
- Harness these technologies to accelerate the pace of discovery in science and engineering, strengthen our national security, and transform teaching and learning.
- Expand the workforce needed to develop and use Big Data technologies.29

The broad sector outlays were in scientific research, health care, defence, energy and geology. Likewise, the UK government announced a cumulative new expenditure of GBP 73 million for Big Data research and development from 2014 to 2017 and estimated that this will create 58,000 jobs in the UK in that period, contributing GBP 216 billion to the economy.30 In the UK, the sectors with funding for Big Data research include the environmental sciences and arts and humanities.

THE INDIAN SCENARIO

The Department of Science and Technology (DST) has announced a Big Data Initiative (BDI) that specifically distinguishes between BDA and Big Data. The initiative is looking to support and fund any aspect of Big Data and BDA, including Science/Technology, Infrastructure, Search/Mining, Security/Privacy and broader Applications. While the BDI is still in its early days, it is encouraging to note that the DST seems to support a broad mandate—from ‘Mining Indian Tweets to Understanding Food Price Crises’ to ‘Population Migration’.

OPTIMIZING BIG DATA EXPENDITURE IN INDIA

The UID will create a unique key for every Indian. This will be similar to the Social Security Number (SSN) in the US, which is used to help the government and corporations in most aspects of their interaction with an individual—from credit ratings, loan applications and fraud to healthcare records, doctor visits and insurance claims. While in a country as diverse and populated as India, implementing the UID is a complex challenge, its real benefits will only be visible when there is near complete coverage.

Food Security

Nutrition deficiency, especially in the young, is one of the leading concerns facing India, where 29.5 per cent of the population lives below the poverty line.31 While the fast-growing population definitely contributes to this problem, there are other factors affecting food security, like agricultural productivity, distribution and weather. Advanced systems that make use of BDA can not only help predict the weather, but also recommend improvements to the distribution system and merge existing information to create social safety nets to combat hunger at a fraction of the cost now incurred.

- **Agricultural Productivity**: Precision agriculture is the ability to use maps, GPS enabled tractors, automated irrigation systems, satellite and drone images, etc., to map the current crop yield rates and ascertain optimal cropping patterns by utilizing data points on soil quality, weather forecasts, fertilizers, water availability, etc. Moreover, techniques for water management and ground water monitoring are used in conjunction to determine water availability through ground and rainwater harvesting. The US has been an early mover in mechanized agriculture production and despite a surprisingly low technology penetration in certain parts of that country, larger corporations are already using BDA to help implement precision agriculture. Intel has taken a lead in California to encourage farmers to collect and share data to help create solutions that transcend the individual farmer and create positive externalities for all.32 While these methods have primarily been the domain of the larger farms in the US, they are also being adopted by smaller farms and co-ops.
US, government intervention at the village level in India can use this technology for productivity analysis and tie it to price data from the mandis.

- **Distribution:** Amartya Sen in his seminal book, *Poverty and Famine,* uses empirical evidence to argue that famines in India rarely happen due to a material shortage of food. They are the effects of broader fundamental problems like unemployment, low wages and most importantly, poor food distribution systems. Thus, improving distribution systems are as crucial to food security in India as is agricultural productivity. Supply chain analytics can make use of satellite maps, GIS data, traffic data, data from government and private warehouses on storage amounts, current production and demand imbalance and overall financial conditions of the local population to predict primary food needs for every locality or village. Notably, the government ‘already’ has most of the data needed to do this analysis. To enable analysis and subsequently, prediction, the data needs to be collated using large-scale data distribution platforms. While professional frameworks like Amazon Web Services or Windows Azure will make this process prohibitively expensive, the cost of analysing this multiple source data can be minimal if the government uses a state or local level Map Reduce framework to access government office computers post work hours for setting up a Cloud Computing structure.

- **Weather:** By making use of the latest analytics techniques on satellite data, forecasting short-term weather patterns has significantly improved in accuracy in the last few years. However, Hu and Skaggs also emphasize regional differences in the quality of weather forecasting in the US. To overcome these potential regional problems, the Indian government needs to implement robust weather monitoring and data collection systems that will ensure high accuracy forecasts across the country. This weather data can then be merged with demand, price and soil quality information to help farmers optimize crop choices.

**Fiscal Revenues and Expenses**

The UK public sector is estimated to lose GBP 20.6 billion every year due to fraud, while the US government lost more than USD 115 billion in improper payments in 2011 alone. While there is no annual tax evasion estimates for India, the ‘black economy’ is estimated to be as large as the regulated sector. A news report from November 2013 states that the Government of India plans to use Big Data and Analytics to improve tax collection. While there were no official statements verifying this, the article quotes Infosys Vice President and India Business Head, Raghu Cavale: ‘Government is planning to use Analytics to increase its revenue base.’ The mandatory use of PAN cards for all banking and credit card accounts as well as tax returns, provides the government with an unique key to help track an individual’s expenses in relation to their declared income and taxes. An enterprise approach to tax evasion and fraud in the government sector can help increase the size of the tax net, especially in a country like India. Data like income, expenses or taxes are all kept in separate silos by the government and this allows defaulters to reduce their tax burdens by under reporting income. If the government, like credit ratings agencies or banks, uses a broad based approach and utilizes Big Data to bring these various data points under a single umbrella, it will be able to significantly improve its tax net. The cash economy is a significant impediment in this attempt to collate income and expenditure data in its entirety. However, large purchases by households like cars, homes or foreign travels can be tracked and tallied with an individual’s declared income levels. Regression techniques like logistic regression or decision tree techniques like CART or CHAID analysis can then be used on these combined data sets for fraud detection.

**Defence and Internal Security**

India is close to the top of the list in the Global Terrorism Index. Between 2002 and 2011, India was ranked 4th in

---

33 Sen, A. (1990). *Poverty and famines: An essay on entitlement and deprivation.* UK: Oxford University Press.

34 Hu, Q. S., & Skaggs, K. (2009). Accuracy of 6–10 day precipitation forecasts and its improvement in the past six year. *Science and Technology Infusion Climate Bulletin*, NOAA’s National Weather Service, 7th NOAA Annual Climate Prediction Application Science Workshop, Norman, OK, March 24–27. Retrieved from http://www.nws.noaa.gov/ost/climate/STIP/RServices/huq_032509.htm

35 PTI (2013). Government plans to use Big Data analytics for tax collections: Infosys. *Economic Times.* Retrieved from http://articles.economictimes.indiatimes.com/2013-11-04/news/43658577_1_lakh-crore-big-data-tax-collections

36 National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2013). Global terrorism database. Retrieved 1 December 2014, from http://www.start.umd.edu/gtd
the list of countries with maximum number of deaths from terror incidents—significantly ahead of recognized terrorism-affected countries like Somalia (Rank 6) and Israel (Rank 20). A recent book *Indian Mujahideen: Computational Analysis and Public Policy* uses mathematical models to study Indian terrorism, however, this has not yet translated into policy prescriptions.

The US Department of Defense (DOD) has been on the cutting edge of defence research since the Second World War. From techniques like Game Theory and Modern Optimization Techniques to technologies like Laser and Sonar, the US defence establishment has been a leader in strategic defence techniques. In the last few years, the DOD has been ‘placing a big bet on Big Data’, investing USD 250 million annually across departments in this area. A prime focus of the DOD’s spending is on analytical and scientific techniques, which can also be used for internal and external defence. The ADAMS (Anomaly Detection at Multiple Scales) programme is developing technologies for detecting data anomalies and cues in massive datasets, which potentially can help in fraud detection, but the DOD is exploiting it for insider threat detection. The DOD has also announced programmes on Machine Reading, Resilient Clouds and Encrypted Data.

India, too, should rely on state-of-the-art research and enable its defence establishment to ethically collate data to analyse and predict potential terror incidents. Web analytics and text mining in regional languages can be handy tools that can support anti-terrorism undertakings.

**Healthcare**

According to a 2013 World Health Organization report on healthcare systems, India ranks 112th among the 190 countries in the study. It is far behind countries with significantly lower per capita GDP or national growth numbers, stressing the challenges due to inequality, regional imbalances and lack of basic medical facilities. According to IBM, ‘Healthcare organizations are leveraging Big Data technology to capture all of the information about a patient to get a more complete view for insight into care coordination and outcomes-based reimbursement models, population health management, and patient engagement and outreach.’ Big Data can be used in India to significantly reduce costs by collating data about an individual patient to successfully predict and thus, prevent health tragedies. An estimated 80 per cent of the health system costs in the US were from chronically ill patients with diseases such as hypertension, diabetes or heart failure. A UID based public health care system, with direct information interchange between schools, private hospitals, primary health clinics, workplaces, doctors and insurance companies, with tight checks on data security, can provide a relief to the system by early intervention for such diseases. An organized streamlining of the system will ensure that the money saved can then be pumped back into the health care system of the country to help it tackle the bigger issues. Other application include epidemic management, doctor–patient matching, kidney (or organ donation) matching programmes (successfully used in the United States) or free healthcare and medical access for Below Poverty Line (BPL) patients.

**CONCLUSION**

Governments help create and can be the biggest beneficiaries of the data that is generated by a country. In a country like India, with its 1.2 billion people spawning enormous amounts of data every day, there is a unique opportunity to use Big Data Analytics to control the data behemoth and tame it for the country’s benefit. From healthcare to defence and the all-important food security issue, Big Data Analytics can help reduce costs, provide insights and policy prescriptions, and help oversee the implementation of policy and assist in monitoring their effects. While a significant body of knowledge already exists on the subject of Big Data, international research cannot substitute the need for organic, localized investigation. From local language search and text mining to finding patterns in disease propagation, the answers can only come from within the country of interest.

The Government of India’s Big Data Initiative is still in its early days and is certainly a stride in the right direction; however, different sections of the government, other than the Department of Science and

---

38 Subrahmanian, V. S., Mannes, A., Roul, A., & Raghavan, R. K. (2013). *Indian Mujahideen: Computational analysis and public policy*. Springer. Retrieved from http://link.springer.com/book/10.1007%2F978-3-319-02818-7

39 OSTP (2014). Executive Office of the President of the USA.

40 Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2010). The Big Data revolution in healthcare: Accelerating value and innovation. McKinsey Insights & Publications. Retrieved from http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care
Technology need to be an intrinsic part of the process. All decision-making bodies should try to get expert level access to Big Data analytics to help them create public policy. Ministries like Defence, Health, Finance and Agriculture will need to take initiatives and become active participants in this technological revolution.

Cloud Computing and Big Data: The Greenwash Perspective!

Krishnadas Nanath

While there is a misconception that Cloud Computing is still a buzz word, Google Trends and Gartner’s Hype cycle reveal a different picture altogether. A quick comparison of search trends on Cloud Computing and Big Data reveals interesting perspectives. As of December 2014, Big Data overtook Cloud Computing in terms of popularity, news events, and search index (Figure 5). It is interesting to note that during initial stages (2006–2007), Big Data had higher search index than Cloud Computing which repeated itself in 2014. The search trends have a lot to reveal about the two technologies. The reason for longitudinal interest of two technologies transcends the technology itself and highlights the marketing and adoption of these technologies in various industries today. The positive and negative interpretation of this graph for each technology will be discussed separately.

Figure 5: Comparing Cloud Computing and Big Data

It is also important to consider the Hype Cycle report for these technologies. As of 2012, Big Data was in the ‘Technology Trigger’ stage and moving towards ‘Peak of Inflated Expectations’. Technology trigger is the stage when there is a substantial amount of time devoted to research and development and the technology is getting ready to penetrate the market. Peak of inflated expectations, on the other hand, is the next level where there is a lot of interest and media reports about a technology and the awareness level reaches a marginal difference when compared to the previous stage. Cloud Computing, on the other hand, had moved down the curve from ‘Peak of Inflated Expectations’ to ‘Trough of Disillusionment’. The results of hype cycle are also consistent with the Google trends. However, a quick look at the Hype Cycle of 2014 reveals interesting insights again. While both the technologies moved forward in the hype cycle, the predictions for the number of years needed to reach the plateau of productivity was not consistent for Big Data. While it was predicted that Big Data would require 2–5 years (in 2012) to reach the plateau of productivity, this number changed to 5–10 years in 2014. Therefore, combining Google Trends and Gartner hype cycle, there are a few concerns that the associated stakeholders should ponder upon. An overview of progression in Hype Cycle is provided in Table 1.

Table 1: Hype Cycle Progress for Cloud Computing and Big Data from 2012 to 2014

| Technology   | Trigger          | Peak of Inflated Expectation | Trough of Disillusionment |
|--------------|------------------|------------------------------|----------------------------|
| Cloud Computing | X              | X X X X X                  | X X X X X                 |
| Big Data     | X X X X X       | X X X X X                 | X X X X                   |

Note: A five-point scale indicates the length of curve from one stage to another.
The first concern as evident from Table 1 is that Cloud Computing has almost reached the Trough of Disillusionment. This does not necessarily mean that people are losing interest or that the technology is fading away. However, trends of interest from Google also reveal that the interest is declining as of 2014. This is the right time for the industry and associated stakeholders to work collectively for defining standards and best practices in Cloud Computing for it to reach the plateau of productivity.

The second concern is the progress of Big Data on Hype Cycle. The progression along the curve had been quite slow and it still lies in the peak of inflated expectation stage as of 2014. While pace of progression is not a real concern, the actual problem lies in the prediction of Big Data reaching the plateau of productivity which has changed from 2–5 years (in 2012) to 5–10 years in 2014. This indicates a lack of maturity and implementation standards in the technology and a lot of vagueness about the definition of Big Data itself.

This article takes up each technology to understand the current problems in the industry and implementation from a management perspective. It also highlights the practice of several firms associating names like Cloud and Big Data with their products and services, with no real implementation of the same. The question of what qualifies as Cloud Computing is still better answered today, but it is not the same for Big Data.

**‘BUZZ’ WORDS NO MORE?**

While it is true that Big Data and Cloud Computing have made major leaps in technological advancement, the real issue to be addressed is the disassociation of the word ‘Buzz’ with these technologies. Drawing insights from Porter’s strategies of competitive advantage, IT industry is no alien to the race for market share. Cloud Computing has been perceived as a highly disruptive technology and the future is envisaged to be a platform where computation moves from local computers to centralized facilities operated by a third party. This section attempts to discuss the technologies from a perspective that calls for defining more standards in the industry.

**Race to be a ‘Cloud’ Company/Product: The Problem**

Often the association of word ‘Cloud’ adds value to a product/company and portrays an image of being updated with the recent trends in industry. However, an end-user hardly takes any effort to understand the technical implementation of Cloud or even to find out if a service is Cloud based or not. Several companies face the pressure of entering into Cloud market to keep up to the competition and on many occasions they end up with associating just the name ‘Cloud’ on a simple Client/Service model. This concept of associating ‘Cloud’ with all possible services without really implementing technologies like virtualization came to be known as ‘Cloud Washing’.

Cloud Washing is derived from a word well associated with sustainability, named Green Washing. Famous hotel chains would recommend their customers to put their used towels on a shelf to avoid washing. While the intentions were to save costs and increase revenues, it was associated with sustainability and green. Similarly, the idea in Cloud Washing is to associate the word ‘Cloud’ with all possible services irrespective of the underlying technologies and properties of Cloud Computing. The National Institute of Standards and Technology (NIST) specifies several properties of Cloud Computing that enable ubiquitous, convenient, on-demand network access to a shared pool of configurable resources that can be rapidly provisioned. The association of word ‘Cloud’ with several applications and services is quite easy, but not many succeed in providing associated properties.

Mell and Grance (2011) defined three service models initially—Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). However, with the growth of technology and technical offerings, the term ‘as a Service’ also started getting associated with all possible offerings internet had to make. Examples include Backend as a Service (BaaS), Security as a Service (SecaaS), Data as a Service (DaaS)

---

41 Rimal, B. P., Choi, E., & Lumb, I. (2009). A taxonomy and survey of cloud computing systems. In INC, IMS and IDC, August 2009. NCM’09. Fifth International Joint Conference, 44–51.
42 Foster, I., Zhao, Y., Raicu, I., & Lu, S. (2008). Cloud computing and grid computing 360-degree compared. In Grid Computing Environments Workshop. GCE’08, 1–10.
43 Yang, H., & Tate, M. (2009). Where are we at with cloud computing? A descriptive literature review, 20th Australasian Conference on Information Systems, Melbourne, 2–4 December. Retrieved from http://www.academia.edu/1325008/
44 Mell, P., & Grance, T. (2011). The NIST definition of cloud computing. National Institute of Standards and Technology, 53(6), 50.
and others. Ramifications of associating ‘as a Service’ is not limited to transcending NIST categories, but the real problem lies in the fact that many of the services are just names associated with Cloud instead of the computing architecture.

Cloud Washing definitely tends to hurt the industry and future of technology in the years to come. It not only projects Cloud to be a marketing gimmick but also creates confusion about what Cloud Computing really is. Assuming Cloud Washing will be taken care of, there are still many problems that question the technology reaching ‘Plateau of productivity’ in the hype cycle. These problems are mostly from a strategic perspective and question the real use of Cloud Computing in the years to come. Each problem is discussed in detail.

The first problem is the use of Cloud by start-ups or young companies that face a resource crunch and want to pay exactly for the amount of resources they consume. It has been suggested by Cloud service providers that Cloud Computing implementation is beneficial for young companies. However, from a strategic perspective, these companies are losing the technical know-how of maintaining their own servers in the long run if they rely on Cloud Computing. Five to six years down the lane, if the firm decides to switch back to conventional computing (maintaining own servers), the effort would be taxing and will involve financial resources. In order to avoid the cost, firms maintain a mix of own servers and Cloud services. The effectiveness of this strategy from a profitability perspective is extremely doubtful.

The second problem is the migration of data from Cloud to conventional servers when a company decides to discontinue Cloud Computing services. The Net Present Value (NPV) calculations for Cloud Computing should also consider the fact that these services are not for lifetime. Hence, firms should factor in the cost of discounting Cloud Computing services if real profitability is to be assessed. Further, Cloud service providers assure the destruction of data, but there are no means to check the validity of the claim. This again raises doubts on the potential of Cloud Computing to mature as a technology in the long run.

The third problem is the information available for decision-making in Cloud Computing. Though many firms comprehend the need and face the pressure of Cloud Computing adoption, the cost–benefit calculations model and standards are scarce and hence there is always an uncertainty in decision-making. Companies having medium-sized computing infrastructure have a major decision to make—to be a Cloud Computing service provider or to adopt Cloud Computing as a customer. Each decision has its own ramifications on the business of the firm. Though there are several other problems like security, reliability, data risks and others, most of them have been addressed by the industry. Moreover, these problems are from a technical perspective while this article has its focus on management issues particularly from services perspective.

Race for Big Data Implementation: Are Firms Really Doing It?

It was reported by the CIO of a leading firm in the Middle East: ‘We are racing ahead with recent trends in IT industry and we work a lot on Big Data Analytics.’ However, when explored deeper about the techniques, he reported simple Business Intelligence (BI) methods that were used long before the word Big Data was coined. While Big Data has caught the attention of several data scientists in the world, it has definitely been overused and overstated several times. Many firms use simple data mining and BI tools and claim to be in the race of Big Data.

Many people stick to the question of ‘How big should the data be to qualify as Big Data?’ which is a wrong question to ask. While a basic census data of more than 5 million points can fail to qualify as Big Data, a small sample set of 100 people with their genetic information could serve as Big Data. The former example (5 million points) might take just a few seconds to run basic computations like median age, it might take years to slice and dice the latter example and view all the slices. So the real question here is—Is the data wide enough to qualify as Big Data?

Woods reported a rise of ‘fake Big Data’ products after analysing the Big Data services and products in the Strata and Hadoop World conference. The concept of fake Big Data products is similar to that of Cloud Washing where the adjective gets associated with products and services with the hope that the world will take

---

45 Retrieved from http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp

46 Woods, D. (2014). How to identify fake Big Data products. Forbes Technology Review. Retrieved from http://www.forbes.com/sites/danwoods/2014/10/15/how-to-identify-fake-big-data-products/
interest. From a marketing perspective, it has been a popular trend to gain from the updraft of a powerful trend. However, with the rise of Big Data washing will come many problems that will question the maturity and adoption of Big Data.

In a survey conducted by Luth Research for Platfora, it was reported that 55 per cent of the respondents felt that small data solutions were repackaged as Big Data. DataRPM, for example, describes itself as a Big Data company and it solves the top of the funnel BI problem using natural language and semantic modelling of data. The only relationship with Big Data as described in Woods is the possibility of accessing queries against Big Data repositories from this product (also possible with spreadsheet). Hence, this world is not short of fake associations of trending technologies with traditional execution, but this needs to be addressed before the bubble gets busted.

‘To me, an analysis becomes a Big Data analysis when new, massive data sets are included and dots are connected to know more about patterns and outcomes,’ said Ben Werther, CEO and Founder of Platfora. ‘When you combine the usually distinct silos of customer interactions, transactions, and machine data, you are in the realm of Big Data. We think that the critical challenge is making it possible for every business analyst to ask questions that matter right away without an IT bottleneck’.

**TOWARDS A SOLUTION: SUSTAINING TECHNOLOGIES**

While the problems of Big Data and Cloud Computing from a marketing and management perspective is clear from the previous sections, it is important to discuss the solutions related to these technologies. Several promising technologies have faded in the past and the reasons go beyond technical failures. Understanding technologies from a management and implementation perspective helps them to sustain in the long run. One of the most important solutions for the sustenance of emerging technologies like Cloud Computing and Big Data is the strengthening of industry standards. Starting from defining the technology to sustaining its usage in the industry, standards play an important role in the use of technology. While it would be a bit early to set standards in Big Data, Cloud Computing definitely needs collaboration from industry partners to set technology standards. Big Data needs more work from IT Giants like IBM, Google, Microsoft and Oracle to define the technology first and enable qualifiers to name a product/service as Big Data.

It is well known that the benefits of Cloud Computing and the standard processes that emerge from the technology are not tailored for all organizations. There are several processes and recommendations available, but most of them are simply the guidelines. Several Cloud players should collaborate and develop a clear method for each user to yield the best possible results in Cloud. There is a need for solid vision, and research would then suggest that Cloud potential will surpass IT expectation eventually. Such technologies demand good understanding of business coupled with multi-domain expertise allowing companies to design, build and operate efficient IT. These technologies must seamlessly integrate with the existing environment of the organizations and leverage rigorous automation to deliver value.

---

47 Ibid.
48 IDC (2009). IT cloud services forecast: 2009–2013. IDC announcement.
49 Harms, R., & Yamartino, M. (2012). The economics of a cloud. Retrieved from http://www.microsoft.com/en-us/news/press-kits/cloud/docs/the-economics-of-the-cloud.pdf
50 Yang, H., & Tate, M. (2009). Where are we at with cloud computing? A descriptive literature review, 20th Australasian Conference on Information Systems, Melbourne, 2–4 December. Retrieved from http://www.academia.edu/1325008/
51 Kerlesz, A., Kecskemeti, G., & Brandic, I. (2009). An SLA-based resource virtualization approach for on-demand service provision. In Proceedings of the 3rd international workshop on virtualization technologies in distributed computing.
Big Data and Consumer Privacy

Pulak Ghosh and Janakiraman Moorthy

Consumer privacy, collection of personal data, analysis and utilization are long standing concerns of marketers. These concerns form an integral part of the technologies that have a growing influence on the society. More than a century ago, Warren and Brandeis, legal scholars in their work on potential impacts of the use of new photographic technologies on media business, made an interesting observation: ‘What is whispered in the closet shall be proclaimed from the house-tops.’52 These concerns are much more widespread and significant, with the ubiquitous deployment of technologies such as the Internet, mobile technology, smart phones, Google Glass and convergence. Advancements in Big Data pose a number of new challenges in terms of ways and means of data collection, analysis, interpretation and its use. Let us look at some of the cases where consumers are outraged.

Dana Mattioli53 reported in Wall Street Journal that Orbitz.com an online travel search engine was displaying different, and costlier hotel search results to online consumers who were using their Apple devices when compared to their PCs. Consumer presumption is that the search engine provides best options based on the parameters specified by them. What Orbitz pushes to the user is a combination of what the user has entered and its own information generated by device analytics. Orbitz could profile the customers based on the data that may not necessarily be permitted by the customer for its use. Orbitz CEO, Barney Herford defends this process, by saying that they found Mac users to be 40 per cent more likely to choose 4–5 star hotels, when compared to PC users.

Charles Duhigg,54 a staff writer for The Times and author of The power of habit: Why we do what we do in life and business, vividly describes the development, deployment and aftermath of predictive analytics in Target Retail Stores. Andrew Pole, Statistician at Target was posed with a question by his Marketing Department, ‘If we want to figure out if a customer is pregnant, even if she doesn’t want us to know, can we do that?’ The inquiry started with that question, Pole and colleagues ran several tests and models and developed a method to assign a predictive score for pregnancy based on the purchase patterns of the products. Using the scores, they started targeted communication and promotions such as coupon distribution. Duhigg narratives how an angry parent approached the Target store Manager at Minneapolis and questioned the mailer that was sent to his daughter. He questioned the Manager about Target’s intentions, ‘She’s still in high school, and you’re sending her coupons for baby clothes and cribs? Are you trying to encourage her to get pregnant?’ The Store Manager was perplexed about the claim as he was caught unawares. He could only confirm that the mailer was from Target and that it did carry advertisements for maternity clothing, nursery furniture, etc.

Netflix, a known on-demand and Internet movie streaming company, also delivers DVDs on mail. Netflix announced a predictive analytics contest for recommending movies based on movie rating and customer profile. It released anonymized data sets to the contestants. Two researchers from the University of Texas identified individual users by developing a method by matching the data sets with film ratings on the Internet Movie Database. In a similar manner, an MIT graduate student, Latanya Sweeney using some simple techniques with voter list and data released by Massachusetts Group Insurance Commission meant for improving health care and controlling cost, identified the Massachusetts Governor William Weld.

At the outset, revealing more and more personal data seems to be unavoidable. With several fold increase in internet access, wireless technologies, mobile technologies, smart phones, convergence, the Internet of things, Facebook, Twitter, Google etc., people unknowingly and unintentionally are communicating personal data to someone else. For marketers Big Data is a powerful weapon for capturing consumer data directly, indirectly, unobtrusively, with and without permission.

---

52 Warren, S. D., & Louis, D. B. (1890). The right to privacy. Harvard Law Review, 4(5), 193–220.
53 Dana, M. (2012). On Orbitz, Mac users steered to pricier hotels. Wall Street Journal. Retrieved 10 January 2015, from http://www.wsj.com/articles/SB100014240527023044586045866045774888226673253882
54 Duhigg, C. (2104). The power of habit: Why we do what we do in life and business. New York: Penguin Random House.
and participation. It provides enormous potential to precisely and efficiently identify behaviours, behavioural changes and target them at the individual level. Combination of Big Data and mobile devices can create new offers that are spatially and temporally very relevant to the consumers. Such an unequal power disparity between customers and corporations can create a constant battle between them. Legal, ethical and privacy concerns are very imminent, complex and different from the earlier days.

As a response to some of these issues on the privacy part of the use of data, the United Nations Secretary-General’s High-Level Panel on the Post-2015 Development agenda has called for a data revolution for improved accountability and decision-making, and to meet the challenges of measuring sustainable development. As a way forward, the UN has taken an initiative called ‘Global Pulse’ which is a flagship innovation initiative of the United Nations Secretary-General on Big Data. Its vision is to create an environment in which Big Data is being used safely and responsibly for the public good. Keeping this in mind, the UN Global Pulse has created a high-level Global Pulse Data Privacy Advisory Group by bringing together experts from public and private sectors, academia and civil society, as a forum to engage in a continuous dialogue on critical topics related to data protection and privacy with the objective of unearthing precedents, good practices, and strengthen the overall understanding of how privacy protected analysis of Big Data can contribute to sustainable development and humanitarian action.

Keeping the above issues in mind, the advisory board recommended a three-fold action: (a) individual members lend expertise to inform the development of guidelines and practices that mitigate risks associated with privacy in Big Data analytics, while preserving utility for global development; (b) how Big Data Analytics can be used for responsible social value creation; and (c) participating in an ongoing privacy dialogue, providing feedback on proposed approaches and engaging in a privacy outreach campaign.

**BIG DATA MARKETING—PRIVACY ISSUES**

Market researchers face a number of ethical and privacy challenges with the emergence of Big Data.⁵⁵ What they might have considered earlier as free from privacy concerns can be combined with other sources of information that may reveal the identity of groups or individuals. From our earlier example of Netflix and data from Massachusetts Insurance combined with other data sets helped the researchers to identify some individuals. Though the number of individuals in a large dataset is very negligible, given the rate to new technologies and powerful algorithms, it would be a big threat for the market research community. Acquisti et al.⁵⁶ provide another example of a different nature, from experimentation. Researchers could use facial recognition algorithms to analyse publicly available information with photographs from social network sites to identify people in a dating site who were disguised. Such attempts create unintended consequences. Customer trust on the researchers and organizations get eroded. Organizations may not be able to predict the kind of uses or misuses that can happen with the data in the future.

Data security is another major concern. Information security breaches are not totally avoidable as it can’t be devoid of any access. There are several cases of Internet security breaches. The retail industry has reported data thefts, and credit card frauds are still prevalent. Given that data is being generated through multiple and varied sources, it may be difficult to visualize what is going to be a vulnerable point. Often the weakest link is human.

Automation of data collection through embedded systems and interconnectivity of data sets pose another kind of threat. A lot of data gets generated without the consent of the consumers. Usage of cookies on the computer may be by permission but its activities are not necessarily known to the novice user. Search engines, social networks, and online retail stores can capture almost all online activities of a user without getting specific consent. Use of RFID chips in loyalty cards, CCTV cameras, mobile phones and the IoT can help the marketer to accumulate data automatically and unobtrusively without intervening in the activities of the customer. Such data gathering methods lead to large volumes of high velocity data. Real-time data analysis can be used for responding to the customer when they are in need. However, customer consent may not be available for specific data or their activities. Hence, there is an ethical issue. How far organizations can go to collect data and analyse it, and the purposes for which the results can be deployed, is a question.

---

⁵⁵ Nunan, D., & MariaLaura, D. (2013). Market research: The ethics of Big Data. *International Journal of Market Research, 55*(4), 2–13.

⁵⁶ Acquisti, A., Gross, R., & Stitzman, F. (2011). Faces of Facebook: Privacy in the age of augmented reality. Presented at BlackHat Conference, Los Vegas, 4 August 2011.
There is a considerable drop in data storage costs and an increase in value for historical data. Data can now be stored for longer periods of time. Without customer’s control on it, there is a possibility that it may be put to unacceptable purposes. Data remains on social media sites long after the user ceases to use the account.

Currently, organizations are analysing only a small portion of the data that they generate due to limitations. This constraint may disappear in the future and the context of data analysis would change. Availability of such diverse sources of data and analytical capabilities may give rise to new ways of intruding into the lives of customers.

Several governing principles are used for shaping the privacy regulations in different countries. Fair Information Practices are used by the United States, some important governing principles being Notice/Awareness, Choice/Consent and Access/Participation.57

- ‘Notice/Awareness principle’ is intended to allow the person who is sharing data to make informed choices for allowing for data collection, use of personal information and either consent or not to consent for collection and use of data. In case of Big Data, corporations and researchers may face a problem defining exactly what data is likely to be collected and how is it going to be used in future. For example, the CCTV footage of a retail outlet may capture a plethora of information which may not be possible to comprehend at this point of time. The possibility of long-term storage and a combination of other sources of data may create an opportunity for someone else, which may have unintended and negative consequences.

- ‘Choice/Consent’ relates to obtaining permission from the person for collecting data, especially personal information. Most of the regulators consider ‘Do Not Collect’, ‘Do Not Track’, ‘Do Not Call’ and ‘Do Not Target’ as integral parts of customer consent. In the Big Data ecosystem, corporations and data brokers are able to identify the users and predict their preferences and choices without having any data that traditionally consented for sharing the data by the customer. Target Supermarket predictive analytics is an example for such a strategy implementation in the retail industry. Navetta explains with an example how a customer may not understand and foresee the use of personal data in targeting them later. How a customer who has purchased a deep fryer will be targeted by an insurance company later and profiles her as a high risk customer and charges a higher premium. Navetta states, ‘In the world of Big Data, the initial, relatively innocuous data disclosure (that was consented to), could suddenly serve as the basis to deny a person healthcare (or result in higher healthcare rates).’

- ‘Access/Participation’ principle is related to customers’ ability to have access to their personal data to verify its authenticity, accuracy and completeness. The customer has the access to correct inaccurate information about them and add additional information to make it complete. Big Data environment poses new challenges in implementing access/participation principle. Several organizations are involved in data collection, storage, processing and final utilization. Many of them may not be visible to the customers. For example, for increasing the transparency, American Federal Trade Commission advises data brokers to create a centralized website, disclose their identity to the customers, share details of the data collection and use of data from customers, and explain the nature of access they provide to the customers.

Anonymization and de-identification are suggested to tackle the problem of privacy in Big Data environment. The cases of Netflix and Massachusetts Insurance are noteworthy examples where researchers could re-identify some of the individuals’ identity. Though by number such revelations are very few, it indicates the kind of problem that may occur in the future. Robust de-identification algorithms are being developed and reported in the academic world and are yet to be tested in practice.

As Big Data is disruptive in nature, new approaches are being attempted. Alex (Sandy) Pentland of MIT Media Lab suggests that Big Data practitioners can adopt the ‘New Deal’58 to avoid customer privacy issues.

---

57 Navetta, D. (2014). Legal implications of Big Data. The Computer & Internet Lawyer, 31(1), 1–5.

58 The ‘New Deal’ is a series of domestic initiatives implemented in 1930s in USA for bringing transformation. Professor Alex and colleagues of MIT Media Lab realized the consequences of power of using data in the Big Data era. They initiated a ‘New Deal on Data’ Program in World Economic Forum. In the marketing context data is mostly about people, and privacy, data ownership, and data control issues are likely to be of different order in the future. As Alex said, ‘….imagine using Big Data to make a world that is incredibly invasive, incredibly ‘Big Brother’. The experiment in Italy is to create a societal change to make data and its utilization more transparent by making people involved in the process at various stages.
issues. Alex and his colleagues set up safe-harbour areas in Europe on experimental basis. In Trento, Italy, a number of households are living with the ‘New Deal’ on data. The participants are provided with notification and control of data generated about them. Transparency and trust are important pillars of such arrangements. Customers should develop confidence about the entire machinery that their personal data sharing will be better for the economy, and will not worsen either immediately or in the future.

CONCLUSION

Big Data has already dawned upon us. More and more organizations are likely to gain competence in dealing with Big Data revolution. Business organizations are likely to take lead in adopting Big Data initiatives. Large corporations such as Amazon, Google and Facebook are already in the big league. The trend of generating multiple sources of data, building data warehouses, attempts to unlock the potential of data using business analytics and visualization trends are only likely to grow over time. Given the nature of complexity, even innovative corporations may not be able to imagine how data will be put to use in the future, who is going to use, what are the implications of utilization, etc.; it is rather important to simultaneously introspect unbiasedly about potential ethical and privacy related issues. It is important to develop systems and processes to avoid such issues well in advance.

CONCLUDING REMARKS

Janakiraman Moorthy

Kesten Green and Scott Armstrong, in the forecasting context demonstrated that simpler tools can just do the job, than complex ones. Jeanne Ross et al. argued strongly that lot of small data applications are sufficient enough for day-to-day decision-making. They caution that a culture of evidence based decision-making and management is required before shifting towards Big Data deployment. It is an important guiding principle that a Data Scientist has to keep in mind. There are cautions about how it is going to emerge or be yet another hype. ‘What no one seems to be saying—or recognizing—is that the death of Big Data is imminent. Big Data’s demise is only years, maybe even months away, and yes, we’ll need a big casket,’ wrote Mike Phelan, in Forbes. However, Gartner believes that, ‘Unlike other Hype Cycles, which are published year after year, we believe it is possible that within two to three years, the ability to address new sources and types, and increasing volumes of data will be ‘table stakes’—part of the cost of entry of playing in the global economy. When the hype goes, so will the Hype Cycle.’ In the same breath, Morgan Stanley foresees that there may be a shift from corporations controlling data towards empowering customers. They claim that, ‘...data management is more about giving customers the technologies they need to store and analyse ‘any’ data set—any type of data, any size of data, for any type of user, and in any timeframe.’ Arindam Banerjee, Tathagata Bandyopadhyay and Prachi Acharya, provide an optimism and hope. They state that, ‘over time, this euphoria will be replaced by a more practical and useful perception of its role as pervasive ether in the organization, supporting and guiding all decision-making, yet not getting in the way of common sense acumen’. The potential possibilities of Big Data for the marketers are limited only by imagination, innovation and ingenuity, as new opportunities arise the associated problems also need to be foreseen and managed to take care of the interests of diverse stakeholders.

The key, of course, is today there are so many pieces and parts and particles that when viewed separately and independently, the mind not only boggles, but also likely curdles. But when all of these seemingly disparate bits of information are put in the hands of a ‘data chef’ using ‘The Cloud’ and the ‘Big Data blasters’, and the ever increasing army of data analysts, they reveal the innermost secrets of any and all consumers in the wink of an eye, or likely faster—Don E Schultz.

59 Berinato, S. (2014). With Big Data comes big responsibility. Harvard Business Review, 92(11), 100–104.
60 Retrieved 31 January 2015 from www.kestencgreen.com/simplerfor.pdf
61 Ross, J.W., Beath, C.M., & Quadragas, A. (2013). You may not need Big Data after all. Harvard Business Review, 91(12), 90-98.
62 Phelan, M. (2012). The death of Big Data. Retrieved 31 January 2015 from http://www.forbes.com/sites/ciocentral/2012/10/04/the-death-of-big-data/
63 Ibid.
64 Banerjee, A., Bandyopadhyay, T., & Acharya, P. (2013). Data analytics: Hyped up aspirations or true potential? Vikalpa, 38(4), 1–11.
65 Schultz, Don E. (2012). Can Big Data do it all? Marketing News, 46(15), 9.
Janakiraman Moorthy is Director and Professor of Marketing at the Institute of Management Technology, Dubai. Earlier he was Professor of Marketing at the IIM Calcutta and IIM Lucknow. He received his PhD from IIM Ahmedabad. His recent research papers were published in the leading scholarly journals such as *Marketing Science, British Food Journal, Journal of Information Technology Case and Application Research, Journal of Database Marketing & Customer Strategy Management*. He has wide experience in the banking and investment industry. He was earlier the Global Research and Project Director of the Institute for Customer Relationship Management, Atlanta USA. He was the Convener of the prestigious CAT Exam 2011.

e-mail: janakiramanmoorthy@gmail.com

Rangin Lahiri is the Practice Director, leading Atos India’s CRM practice while supporting Strategic Business Development for North American Market. With an experience of more than 15 years, Rangin has worked extensively as a Business Consultant in Information Technology (Sales Automation, Marketing & Service Management area), Customer Data Management and CRM Analytics.

e-mail: ranginlahiri@hotmail.com

Neelanjan Biswas is a Business Consultant at Atos with extensive experience in Business Analysis, Risk Management, Analytics, Business Development, Presales, Solution Ideation on Enterprise Data Management, Enterprise Reference Data and Master Data Management area.

e-mail: Neelanjanb@gmail.com

Dipyaman Sanyal is founder and CEO of dono consulting, a boutique quantitative analytics and investment research firm. He has worked for leading financial firms in New York and India including Dow Jones, Blackstone, Sorin Capital (VP, Quantitative Modeling) and Thomson Reuters (Head of Real Estate Analytics). A CFA charter holder and Commonwealth Scholar, Deep has an MS (Applied Economics) from University of Texas, Dallas and an MA (Economics) from Jadavpur University. He is an Adjunct Faculty at IMT, Ghaziabad.

e-mail: deep@dono.in

Jayanthi Ranjan, is Professor in the Information Systems Group of the Institute of Management Technology, Ghaziabad. Her PhD is in the field of data mining from Jamia Millia Islamia Central University, India. She has published five edited books. She is serving on the editorial boards of *International Journal Information Technology, Asian Network for Scientific Information, International Journal of E-CRM, Inter Disciplinary Journal of Information Knowledge and Management*, and is the Associate Editor of *Journal of Applied and Theoretical Information Technology*.

e-mail: jranjan@imt.edu

Krishnadas Nanath is a faculty in the Information Systems Area at Middlesex University, Dubai Campus. He is also lead corporate trainer for Cloud Computing certification courses organized by Global Science and Technology Forum, Singapore. He works in the areas of Green IT, Sustainability and Cloud Computing. He has special interest in Business Analytics, Big Data Management and Quantitative Research. His doctoral work at the Indian Institute of Management Kozhikode (IIM K) was in the area of Green IT and Cloud Computing. Basically, a Computer Science engineer, he had successful professional experiences with Microsoft Research and Honeywell.

e-mail: k.nanath@mdx.ac

Pulak Ghosh is a Professor of Quantitative Methods and Information Systems at the Indian Institute of Management Bangalore (IIMB). Prof. Ghosh serves in the Advisory group of Big Data at the UN Global Pulse, a Big Data initiative by UN. He is also an academic fellow at the Centre for Advanced Financial Research and Learning promoted by the Reserve Bank of India. His main research interests are Big Data, Machine learning, Business analytics, Marketing analytics, Statistics. He has more than 10 years of rich experience in academics and industry.

e-mail: : pulak.ghosh@iimb.ernet.in