RIDING THE "BIG DATA" WAVE FOR INNOVATIVE MARKETING STRATEGIES

Prof.R.Venkataraman
Visiting Faculty, XIME

1. Introduction

The digital web today generates a huge amount of data with increasing complexity and at the same time provides a great opportunity to companies. With the advancements in technology, companies have learnt to handle this new opportunity to their advantage. Big Data today is a hot topic not only in the corporate board room but also in the political arena. According to A.T.Kearney's IT study more than 45% percent of companies have implemented a business intelligence initiative in the form of Big Data analytics. Further it is estimated that more than 90% of Fortune 500 companies currently have at least one Big Data initiative underway. The overall strategy in all these initiatives is to improve performance in traditional segments and create new opportunities to expand product and service offerings. The size of the growing Big Data is estimated at 45 Zeta Bytes by 2020.

1ZB = 10007bytes = 1021bytes = 10000000000000000000000bytes = 1000exabytes = 1billionterabytes. (Wikipedia)

2. Some insights on how big is Big Data

- Every day, 2.5 quintillion (A billion billion: 1 followed by eighteen zeros, 1018) bytes of data are created, with 90 percent of the world's data created in the past two years alone.
- Data production will be 44 times greater in 2020 than in 2009.
- The volume of business data worldwide is expected to double every 1.2 years.
- million customer transactions per hour, stored in databases estimated to contain more than 2.5 petabytes of data (1 PB = 1000000000000000B = 1015bytes = 1000terabytes).

- The enormous data influx is straining IT infrastructures. In a recent survey, 55 percent of executives said that data is slowing down their IT systems.
- Poor data management can cost up to 35 percent of a business's operating revenue.
- RFID (radio frequency ID) systems generate up to 1,000 times the data of conventional bar code systems.
- In just four hours on "black Friday" 2012, Walmart handled 10 million cash register transactions – almost 5,000 items per second.
- United Parcel Service receives on average 39.5 million tracking requests from customers per day.
- VISA processes more than 172,800,000 card transactions each day
- 500 million tweets are sent per day. That's more than 5,700 tweets per second.
- Facebook has more than 1.15 billion active users generating social interaction data
- More than 5 billion people are calling, texting, tweeting and browsing websites on mobile phones.

Source: http://www.atkearney.com/strategic-it/

Gartner was the first firm to name the Big Data phenomenon in the year 2001. Big data has been defined as "challenges and opportunities in data growth". The increasing data growth had three dimensions (Pettey & Goasduff, 2011) namely:

I. Increase involume

II. Increase in velocity

III. Increase in variety

Subsequently in 2012 the Big Data phenomenon was perceived "as 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 optimization" (Beyer & Laney, 2012).

In the industry, however, it had no clear boundaries and only indicated the challenges in data growth which needed to be resolved. In 2011 the McKinsey Global Institute (MGI) in its report gave an in-depth analysis of the potential of Big Data for firms.

This report used an intentionally subjective definition to capture the essence of the difference between data and Big Data thus:

"Big data refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze." (Manyika, Chui, Brown, & Bughin, 2011).

Furthermore the report highlighted the following:

I. Data have swept into every industry and business function and are now an important factor of production

II. Big Data creates value in several ways:

- By creating transparency;
- By enabling experimentation to discover needs,
- By exposing variability,
- By improving performance by way of effectively segmenting populations to customize marketing actions;

The report stated that "Big Data analytics" can thus replace /support human decision making with automated algorithms and innovating new business models, products, and services. It also pointed out that while the use of Big Data will matter across sectors, some sectors are poised for greater gains and there will be a shortage of talent necessary for organizations to take advantage of this new development. Thus organizations which want to ride on this new wave need to

formulate data policies,decide on appropriate technology and techniques,bring forth organizational change and talent,create access to data and put in place a proper organizational structure.

The following two main developments led to the emergence of the Big Data concept:

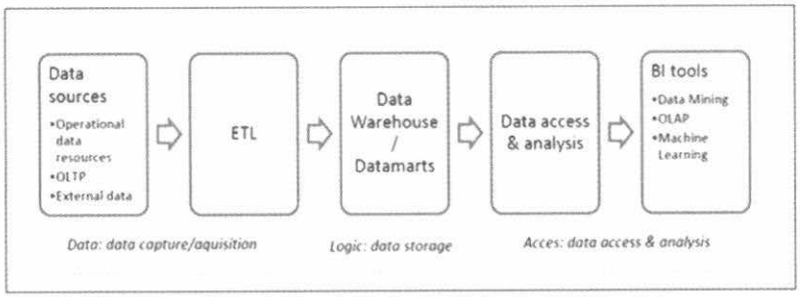
Firstly the amount of data generated increased tremendously during the past few years. Internet exceeded one Exabyte (1 billion giga bytes) per day in 2013 and the first companies facing this explosion were Google and Facebook. In fact these were also the primary force in the development ofa new technology to deal with such extremely large datasets.

The second development is the increasing capacity (Moore's law; Schaller, 1997) and the decreasing price of hardware as a result ofthe commoditization of IT. New open source platforms to analyze large datasets coupled with affordable hardware created the possibility of such larger-scale adoption of the so-called Big Data solutions.

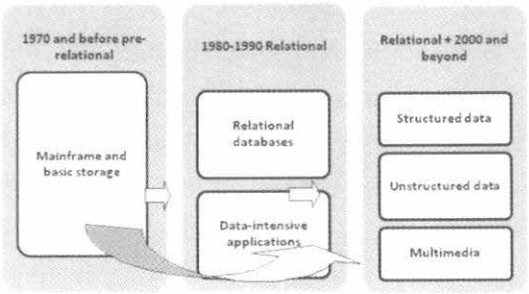
3. Business Intelligence

The term Business Intelligence (BI) was first introduced in the IBM Journal in 1958 stating that the BI system has "the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal" (Luhn, 1958). Luhn described a BI system as: an automatic system being developed to disseminate information to sections of an organization by using data-processing to create profiles for each of the action points of an organization and through incoming and externally generated data. With the development of Decision Support Systems (DSS) starting in the 1960s and the development of computers in the 1980s, Business Intelligence led to the creation of Data Warehouse (DW)or Enterprise Data warehouse (EDW) becoming a central term in BI. But before data can be stored, it needs to be captured from a system which for its part stores data from an operational system.

To be able to have the data stored, it undergoes a process widely known as Extraction Transform and Load (ETL) process. During the ETL process data is extracted from operational systems and it is cleaned to correct missing, inconsistent or invalid values. After extraction, data is transformed into standard formats and stored. Rules are then applied to map the data to enable creationof reports. The final ETL stage loads the data into the storage. After the data has gone through the ETL process, it is stored in a Data Warehouse (DW). In a data warehouse, data is specifically structured for query, analysis and decision support. The data's purpose is to support business decisions, not business operations (Inmon, 1992). Data can also be stored in data marts: small sized data warehouses created by specific departments to facilitate their own decision support (Khan &Quadri, 2012). The data access part of a BI system provides business users with an interface while hiding the technical complexity of the data analysis. The whole process can be schematically represented as below:



The evolution of Business Intelligence and Analysis



Exponential growth in data and evolution of Big Data: (AT Kearney analysis)

Data mining discovers relationships and patterns in data through comparison, characterization, classification, association and cluster analysis (Han, 1997). The main difference between the tools (OLAP) on line analytical processing and data mining is the fact that OLAP operates at the summary level whereas data mining operates at the detail level. We can divide IT systems into transactional (OLTP) and analytical (OLAP). In general, we can assume that OLTP systems provide source data to data warehouses, whereas OLAP systems help to analyze it.

Thus the complete "Business Intelligence & analytics process is often referred to as the techniques, technologies, systems, practices, methodologies and applications that analyze critical business data to assist an enterprise to understand its business and market and make timely business decisions."The field of analytics can be classified into five technical areas which contain emerging analytics research: (big) data analytics, test analytics, web analytics, network analytics and mobile analytics. In the past decade the amount, speed and structure of data has changed tremendously. Applications and research focused mostly on structured data collected by companies through legacy systems and were stored in relational database management systems (RDBMS). Analytical techniques used in this generation were rooted in statistical methods and data mining techniques developed in the 70s and 80s respectively (H. Chen et al., 2012). The rise of semi-structured and unstructured data calls for ad-hoc and one-time extraction, processing, indexing and analytics in the new technological environment. Again with the rapid development of the internet, web-based unstructured content needs to be analyzed. This gets further challenged with the usage of smart phones, tablets and other sensor-based and location-based personalized data generation systems. There is a strong necessity to integrate the first generation BI&A systems with new technology and further these new databases need to be incorporated within the firm's current BI systems. More and more applications for collecting, processing, analyzing and visualizing these large data sets are being developed. While several softwares are available for enabling this, details of these software and their capabilities are not within the scope of this paper.

4. The Importance of Big Data and What Can be Accomplished

Big Data acquisition leads to the obvious question of what to do with it. The vision of the analytics companies is that organizations will be able harness relevant data from the huge volumes of data acquired and analyze it to find answers that enable

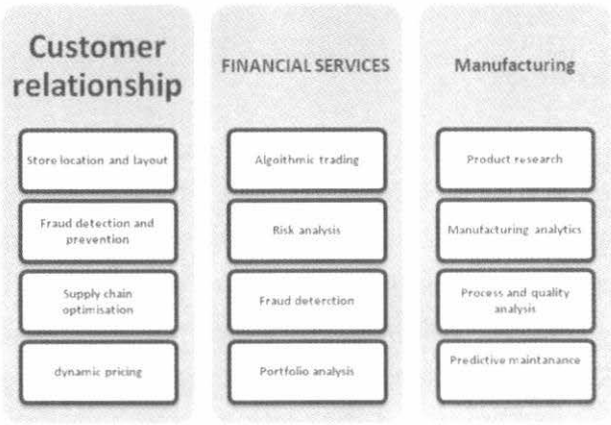
1) Cost reductions,

2) Time reductions,

3) New product development and optimized offerings,

4) Smarter business decision making.

Some of the segments where Big Data plays a very important role are as indicated below:



By combining Big Data and high-powered analytics, it is possible to:

a) Determine root causes of failures, issues and defects in near-real time; effect savings.

b) Route optimization in logistics.

c) Analyze millions of SKUs to determine prices that maximize profit and clear inventory.

d) Generate retail coupons at the point of sale based on the customer's current and past purchases.

e) Send tailored recommendations to mobile devices while customers are in the right area to take advantage of offers.

f) Quickly identify customers who matter the most.

g) Use click stream analysis and data mining to detect fraudulent behavior.

5. Big data and value creation

Big Data analytics supports the creation or improves products, services, markets or channels leading to significant advantages for companies. Research indicates that for many companies the insights drawn from Big Data have already proved profitable and sustainable and have benefited them in three areas, namely,customer intimacy, product innovation and operation.

a) Customer intimacy

Customer data forms the focus for formulation of marketing strategies. Organizations are flooded with huge volumes of customer data obtained from interactive websites, online communities, government and third-party data banks and social-media platforms such as Facebook. The volume is estimated to be nearly 30 billion pieces of content daily. Advanced analytical tools are required to enable faster processing of this data at low cost. With hi-tech software, companies leverage on these converted and structured data sets in order to gain new insights. For instance a retail bank can look at social media activities, gather firsthand feedback and formulate strategies for customer development and performance targets. US retailer Macy's uses Big Data analytics to create customer centric segmentations. Amazon has been doing this for years by displaying an assortment of products as a suggestion to the customers in its "customers who bought this item also bought this" section. In offline advertising, Big Data assists in advertisement placements in determining which TV programmes will deliver the biggest impact for different customer segments.

Organizations using Big Data strategies can also use the information generated in the past to come to understand the trends and developments over the years. An example can be money lending organizations using past cell-phone data to ascertain credit risk. The key strategy is to identify, collate and organize relevant data from many sources or gather off line information.

As a consequence, what is considered proprietary information now can be gleaned from public and private sources and this brings out the privacy issues which will also be discussed in this paper.

b) Product development and innovation

Currently a lot of organizations make efforts to get automatic feedback on their connected equipment to effect product and process improvements or to know more about the consumer and his usage patterns. GE is planning equipment with sensors that will send terabytes of data over the internet back to its product engineers. Due to privacy issues involved, companies normally seek the permission from the user for such a feedback.

More and more companies are resorting to crowd sourcing and other social product innovation techniques that are made possible because of Big Data revolution to transform hundreds of millions of rich tweets, blogs and several unstructured data into insights on products and services that will positively resonate with consumers. These transformed data using sophisticated software and computational machines form the basis of the company's future marketing strategy. This trend has given rise to a number of new technology and analytics firms today; marketing analytics is a very prominent term used in firms for compiling and analyzing transaction data between companies and customers as well as suppliers. Such data owned by them can be used to improve operations, offer new and additional innovative products and services and create revenue streams.

Social network organizations such as Facebook run thousands of experiments daily with different sets of users to create scenarios and analyze them for effecting improvements. E-Commerce companies offer different contents and also dynamic pricing to various customers that fit in appropriately in its target segments.

c) Operations

In a technology intensive supply chain environment micro sensors, radio frequency identification (RFID) readers capture huge data which can be used for improvements and innovations in operational processes and achieve higher revenues and savings in costs. Similarly a retail chain can identify inventory information across its system with the ability to identify overstocks, redistribute the stock and optimize its inventory storing costs as well as effectively avoid locking up its financial resources in the form of inventory.

GE syncs BigData predictive modeling to help airlines that buy GE jet engines to monitor their performance and anticipate maintenance requirements as well as replenishment of the spare parts that will be required. If the data captured indicate that something has started deteriorating, engineers are sent to fix the problem and thus avoid break downs. Only companies which understand that data is an asset can capitalize on these opportunities. For example, when considering the effect of a product promotion, collecting data on competitors' promotions, specifically on substitute products, can reveal how customer preferences have evolved.

6. Creating the organization for Big Data enabled strategies

While it is recognized that tools are available to capture relevant data from the large, diverse data sets to identify complex relationships among the business variables, it is not possible to interpret these extracted data without human intervention. This has brought forth the emergence of a new breed of professionals called data analysts and scientists. These individuals possess advanced statistical and mathematical knowledge coupled with business knowledge in specific domains. They work with the business managers to enable strategic decisions.

With data analysts and scientists on the front lines of the transformation, organizational structures must be redesigned with these expert teams in mind. These organizations need to be restructured to leverage the analytical competencies of these professionals. The climate in the organization must be conducive for them to move from basic and anticipatory to more mature predictive analysis. The basic analytics provides a historical view of the performance with respect to what

happened, when, where and how many times. On the contrary, anticipatory analytics which is the order of the day identifies unique drivers, root causes and sensitivities when implementing strategies. It is this form of analytics that ultimately predict what is likely to happen and help the organizations to be proactive.

7. Leveraging Big data in Marketing

The feasibility of consumer analysis as well as value creation through consumer intimacy has already been discussed in this paper. Hence by combining Big Data with an integrated marketing management strategy, marketing organizations can make a substantial impact in these key areas:

a) Customer engagement. Big Data can deliver insight into not just who our customers are, but where they are, what they want, how they want to be contacted and when.

b) Customer retention and loyalty. Big Data can help in discovering what influences customer loyalty and what keeps them coming back again and again. Thus effective churn management strategies can be formulated.

c) Marketing optimization/performance. Big Data can determine the optimal marketing spend across multiple channels as well as continuously optimize marketing programmes through testing, measurement and analysis of the structured data.

For effective marketing in this Big Data revolution, three of the biggest challenges are:

a) Knowing what data to gather: It is data and data everywhere and hence it has to be the right data.

b) Knowing which analytical tools to use: Today the business competition with more and more technology has become very dynamic and intense. As the volume of Big Data grows, the time available for making decisions and acting on them is shrinking. Analytical tools can help to aggregate and analyze data, provided the right tools are used.

c) Knowing how to go from data to insight to impact: The challenge is how to turn the data sets into meaningful insights. Further competency is needed to use these insights to make a positive impact on the organization's marketing programmes.

Some of the marketing strategies with Big Data are listed below:

Customer Sentiment analysis

In the current day business scenario companies and customers interact actively on the social network. Using data mining tools, marketers can analyze what customers think of their products and services as well as that of competitors. They can then correlate this analysis to sales efforts, product mix, marketing spend, advertising expense, loyalty programmes, market share and customer share. This type of correlation made possible by Big Data analytics is powerful in manipulating the organizations' strategy to influence customer behavior with predictive responses. Eventually companies which track customer sentiments win over customers of companies that fail to do so.

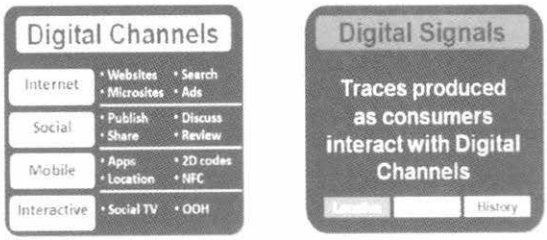
Customer experience analysis

Big Data analysis can track and extract relevant data from data streams and click streams to have a complete perception of the customer experience with products and services and relate this data to demographic, cultural and other variables. This will enable the marketers to detect, measure and improve their strategies to the desired level of customer experience and also recognize the life time value of customers to plan their loyalty programmes. This will also help them to operationalize their up-selling, cross-selling and customer churn mitigation programmes.

Predictive analytics

This is considered as the most effective usage of Big Data to create opportunities. This helps the organization in the development of strategies such as new product introductions, decision to enter new markets, focused segmentation of markets and risk analysis. Marketing executives can now forecast what products to sell to what customer and when it can bundle products according to customer preferences to

increase sale. The customer click stream analysis leads to information on what the customer is looking for. For example, if a person is often surfing the maternity and baby product web sites the marketer can track her click streams and safely predict that the person is pregnant or planning for that and direct its marketing messages and also bundle its products to such a prospect. More and more companies are using Big Data such as comparable and competitor price lists and market responses to decide on pricing for their products and services. Social network data and blogs are being extracted and used for customer perception. The Chief Information Officers identify the segments they need to concentrate on and also can look at the date such interaction has taken place in the competitor's sites. They can also identify the group of people in the social network who have most influence over the others inside these social circles and plan strategies to leverage their sales by addressing these influencers in a unique way for better and effective campaigns.



Thus marketers are able to detect the right digital signals from the various channels to plan their strategies, predict consumption rates and usage patterns. Another important area of focus is how product defects occur or accelerate when used with other products. They're also using this information to take proactive actions and implement corrective measures. In certain situations they are even able to preempt the occurrence of such defects and take preventive actions as has been explained in the case of GE.

These preventive actions are enabled by analyzing customer complaints (tweets, SMS, etc.) along with the volume, trending and responses to those complaints. It is also possible to extract anomalies or patterns by comparing the data to detect signs of product complications. Call Center managers are monitoring customer service call recordings to quickly detect product deficiencies early. Similarly

store security videos are analyzed to plan store configuration based on customer movement patterns. Correlation analysis also plays a key role in prediction and, to cite an example, the correlation of Point of Sale (POS) with weather data helps in understanding how environmental conditions impact product positioning, promotions and sales.

8. Case Studies and application of Big Data in Business:

Big Data at UPS:

UPS has been using Big Data, to capture and track a variety of package movements and transactions as early as the 1980s. The company now tracks data on 16.3 million packages per day for 8.8 million customers, with an average of 39.5 million tracking requests from customers per day. The company stores more than 16 petabytes of data.

Much of its recently acquired Big Data, however, comes from telematics sensors in more than 46,000 vehicles. The data on UPS trucks, for example, includes their speed, direction, braking and drive train performance. The data is not only used to monitor daily performance, but to drive a major redesign of UPS drivers' route structures. This initiative, called ORION (On-Road Integration Optimization and Navigation), is arguably the world's largest operations research project. It also relies heavily on online map data, and will eventually reconfigure a driver's pickup and drop-offs in real time. The project has already led to savings in 2011 of more than 8.4 million gallons of fuel by cutting 85 million miles off daily routes. UPS estimates that saving only one daily mile driver per driver saves the company $30 million, so the overall dollar savings are substantial. The company is also attempting to use data and analytics to optimize the efficiency of its 2,000 aircraft flights per day.(Source: http://www.pressroom.ups.com/Media+Kits/ORION)

Role of Big Data analytics in the Indian General Election:

The Indian General Elections also had another perspective which often does not figure in much of the public media. The complexity of the election process can be understood if one realizes that the whole operation involves 300 parties, 8000

candidates and 800 million voters, 1 million booths served/secured by 20 million officials. The data mix is further complicated with a variety of structured and unstructured information – candidate histories, crime records, declared assets and audacious election manifestos. Mixed with the above is the frenetic activity on the day of results. Live streaming of results: 21000 votes to be counted per second from all corners of the country spanning an area of 1 million square miles.

Thus, election in India is a classic Big Data application and the 2014 general elections were the biggest of them all. While technology may be able to process this data of astronomical proportions, the real challenge is how this information is consumed and understood by a billion people when everything that is happening is in real time. Such large volumes of data was analyzed and converted into meaningful information for the general public to consume easily.

The election timeline had 3 logical phases spread over 2 months: 2 weeks of pre-elections, polling of 5 weeks and post-election phase of 2 weeks. Even before thinking about the imminent elections was the task of assimilating historical data. Data from the 1950s was gathered from multiple PDF files published on the Election Commission of India (ECI) website and followed with the task of cleaning, correction and collation. Synthesizing this data with various other credible data sources completed the painstaking exercise of building an integrated, structured elections data source. This master data source served as the backdrop for 2014 election with over 60 years of election data, analytics and visualization technology used to complete the task. This history was hosted on a web page.(Source:https://gramener.com/election/parliament).

This first step brought about an interesting twist to election reporting enabling people to consume data stories of over 60 years. Soon print media, online and on-air media were discussing these stories since the audience was able to understand simple but powerful information easily – victories with the highest margins, regions which always unseated the incumbent, most persistent yet unsuccessful parties, highest number of candidates contesting for a single seat, counter-intuitive stories like "victory margins of the winner goes up when number of candidates goes up" and so on. For the first time, the electorate in India was exposed to data-based

reporting with hard facts and numbers, which were easy to consume as data stories. The same analytics and visualization techniques were used to report the 5 weeks of polling by providing live facts on voter turnouts, exit poll surveys etc.

Streams of vote-by-vote count and results feeds need to be analyzed, visualized and data stories need to be reported real time on national television. Accuracy, speed and ease of consumption were put to test in front of the entire nation which had reached a crescendo in anticipation. Gramener's analytics and visualization engine, "Gramex" was programmed to arrive at an election results dashboard to meet these heavy expectations. Particular care was taken to enable real-time slice/ dice of data for deeper insights as it streamed in. Again, the challenge was that it had to be novel and farthest from plain reporting of live numbers which had been the norm until then. With plans in place and trial runs completed, the visualization dashboard went live on the morning of 16th May, 2014 – the counting day. The concern of the analysts was "how will the people of India receive this? will they like it? as they understand it?"Gramener and CNN were holding their breaths to know the reception of this by India, laced with trepidation – will Gramener technology stand the ultimate performance test on this D-Day? Answer came in the first hour – the visualization dashboard had over a million hits in an hour. Grew to a total of 10 million hits in a span of 12 hours! Social media exploded with responses in humongous proportions to these live visualizations. Common people started consuming and sharing deeper insights real time! The various mass media channels lapped up the insights and reported them across the county. A new era in Indian election reporting was born. It was not only a new era in election reporting, but it was an epitome of collaboration between mass media and technology companies. (Source: http://ibn.gramener.com/live).

Procter & Gamble and innovation:

P & G is reshaping its innovation through "data modeling, simulation, and other digital tools". P&G created "baseline digital-skills inventory of its employees that's tailored to every level of advancement in the organization". Business intelligence (BI) managers have the skills to provide input into business-unit and corporate strategy and business analysts know how to take a hypothesis-based approach

to problem solving and adopt a cross-departmental view of data to identify opportunities firm-wide. Similarly, data analysts work with unstructured data in developing complex statistical models, and data managers redesign IT architecture so that it provides a company-wide view, incorporating larger data manipulations (of unstructured data) and modeling activities into their decision-making process.

Tesco PLC and performance efficiency:

The supermarket chain collected 70 million refrigerator-related data points coming off its units and fed them into a dedicated data warehouse. Those data points were analyzed to keep better tabs on performance, gauge when the machines might need to be serviced and to do more proactive maintenance to cut down on energy costs.

Wal-Mart Stores Inc. and search.

The mega-retailer's latest search engine for Walmart.com includes semantic data. "Polaris", a platform that was designed in-house, relies on text analysis, machine learning and even synonym mining to produce relevant search results. Wal-Mart Laney said that adding semantic search has improved online shoppers completing a purchase by 10% to 15%. "In Wal-Mart terms, that is billions of dollars".

American Express Co. and business intelligence:

Hindsight reporting and trailing indicators can only take a business so far, AmEx realized. So AmEx started looking for indicators that could really predict loyalty and developed sophisticated predictive models to analyze historical transactions and 115 variables to forecast potential churn. The company believes it can now identify 24% of Australian accounts that will close within the next four months.

Express Scripts Holding Co. and product generation:

Express Scripts which processes pharmaceutical claims realized that those who most need to take their medications were also those who were most likely to forget to take their medications. So they created a new product: Beeping medicine caps and automated phone calls reminding patients it's time to take the next dose.

PredPol Inc. and Predictive analysis:

The Los Angeles and Santa Cruz police departments, a team of educators and a company called Pred Pol have taken an algorithm used to predict earthquakes, tweaked it and started correlating it with Big Data analytics on crime data. The software can thus predict where crimes are likely to occur down to 500 square feet. In LA, there's been a 33% reduction in burglaries and 21% reduction in violent crimes in areas where the software is being used.

Macy's Inc. and real-time pricing: The retailer adjusts pricing in near-real time for 73 million items, based on demand and inventory and using technology from SAS Institute.

(http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell)

How Indian TV show "SatyamevJayate" used Big Data to inspire the world

On 6th May 2012, the first episode of 'Satyamev Jayate' – a television series focused on social issues in India, was aired. The producers who were obviously eager to gauge how viewers rated the show had a team in the background whose members were waiting diligently to execute one of its toughest projects. This team was from Persistent Systems which was given the task of analyzing messages generated from the social media. The show's producers wanted to analyze the messages that they would receive on the social media to not only improve the future episodes, but also use the data to push for improved governance. For Persistent Systems, which had done many Big Data projects earlier, this project was unique as the scale and the type of data that would be received was difficult to comprehend. They had no precedent in the industry which could help them build a system. The data could in different formats from a variety of different sources such as Facebook, Twitter, websites, SMS polls and phone voice messages. This data needed to be analyzed immediately to understand what type of people were liking, and the impact that the show had at an individual and society level. What made all this much more complex was the fact that there was no dry run. All Persistent Systems could do is gauge the type of response that could be expected. An intense marketing activity and an actor in the form of Aamir Khan hosting the show meant that the show was

expected to have a huge viewership and participation. Additionally, as the show was based on social issues, it was expected that a large amount of information would be meant to be collected. While it had no operating guidelines, Persistent designed a system it believed was flexible enough to handle the load. The huge popularity of the show indeed surprised everybody when the flood of data began flowing in from all sources. The first episode on female foeticide gathered 1.4 million responses from all sources. The data was in different formats – text, audio and video and in different languages that modern India was comfortable with English and Hinglish (Mixture of Hindi and English). As the popularity of the show grew, so did the tweets and the messages from different social networks. More than half million tweets were aggregated for season 1. The traffic was around 40,000 tweets on an average during the 90 minutes of the show and it was observed that the Twitter traffic was higher on Sunday and Monday. A two-phased approach for the Live Analytics was used during the show. The analysts leveraged crowd-sourcing for analysis and, for deeper insights, they built algorithms to filter out the relevant tweets. A software platform was developed by assembling an array of automated tools to parse the data and a user-interface for several analysts to process messages for deep analytics. The result was a cluster-based analysis along with trend, demographics and sentiment analysis for each message. The final step involved a manual check to find the latest and relevant top story. The results were aggregated and then further used for creating visualizations and dashboards. The analysis was done for all the 13 episodes over a thirteen week period.(http://www.informationweek.in/informationweek/news-analysis/175873/indian-tv-satyamev-jayate-inspire-world?utm_source=referrence_article)

9. Risk management in Big Data management

Leveraging Big Data for SMART business must start with an integrated people, process and technology plan which includes the processes to identify and capture the data, the tools to manage the data, and the right-time distribution of that data to the person or interaction where it can be applied for specific purposes and consistent results. Companies must consciously make efforts not to put technology ahead of people and processes.

Further the data scientists and business managers should be conversant enough to understand the relevance and timing of data. The management must make efforts to learn challenges such as data privacy, information security, information distribution, data presentation and data overload. Huge sets of external data are available from government and NGO bodies, social media and commercial services and the project can be implemented smoothly if the personnel have good domain knowledge.

10. Privacy issues in Big Data strategy

Following President Obama's speech on reform of the National Security Agency's bulk meta-data collection, the first major privacy initiative announced by the White House involved key officials across the federal government, including the President's Science Advisor and the President's Council of Advisors on Science and Technology. The deliberations focused on the following areas:

1) What potential harms arise from Big Data collection and how are these risks currently addressed?

(2) What are the legal frameworks currently governing Big Data, and are they adequate?

(3) How could companies and government agencies be transparent in the use of Big Data?

(4) What technical measures could promote the benefits of big data while minimizing the privacy risks?

(6) What future trends concerning Big Data could inform the current debate?

On May 1, 2014, the White House released the Big Data Privacy Report. The President's Council of Advisors on Science and Technology ("PCAST") also released a report on the same day, entitled "Big Data and Privacy: A Technological Perspective."

Data brokers who are large commercial organizations that collect vast data on millions and sometimes hundreds of millions of consumers in order to resell the data or utilize it in targeted marketing campaigns have come under a great deal

of scrutiny from the Federal Trade Commission (FTC). The FTC has promised to promote more transparency in the data broker industry and to give consumers greater control over their individual data. The Senate Commerce Committee has also promulgated a bill entitled The Data Broker Accountability and Transparency Act. This act is designed to provide some broad guidelines for regulating the data broker industry.

While the recent literature on Big Data does not provide evidence of identity fraud or data breaches as far as the legal use of data for commercial purposes are concerned, the various watch dog agencies have also found no evidence of an increase in harm to consumers. However, the data brokers perceive the FTC to be an overprotective steward, often reducing consumer welfare by excessive regulation of information.

Owing to privacy issues, about 49% of the consumers are less willing to share their personal information as they are now aware of the dangers of sharing their personal information and of the security issues even though companies are collecting this information after taking their consent to share their personal information online. Data collection tools and technology have become so innovative in collecting data from imperceptible consumers that the consumers strongly feel that it is an intrusion into their privacy and that they have every right to take legal recourse. Many corporate firms are beginning to address the concerns of the consumers by implementing their own self-regulatory programme in the data gathering process employed on their website and making it known to their customers. By introducing the Big Data process in collecting the personal information of their website visitors and through the public disclosure about the process and technology employed in protecting them, businesses are able to assure the consumer on how their data will be used and the level of protection they will receive in exchange of sharing their personal information. The implementation of such a self-regulating process of data gathering will reduce the hostility among the consumers.

The industry Big Data experts therefore recommend the following best practices:

a) Limit the data collection process to specific parameters without including personal information of the consumer that will reveal their identity.

b) Format the summary records to reveal different levels of information, but excluding the identity of the user where the aggregate data was collected from.

c) Identify the legal and social norms involved in using customer information and take steps to use the data for business decision-making within the realms of reasonableness.

d) Use only relevant information that is extracted from the customers for data analysis.

e) Explain to your customers the steps in protecting their shared information when requesting for their consent to disclosing personal data.

The Indian scenario on Big Data privacy issues

The concept of data privacy and protection is at a nascent stage in India. Framers of the rules have attempted to adopt ideas from jurisdictions which have a long standing and mature data protection regulations. These rules are therefore only a first step and stringent implementation of the law and healthy development of the data privacy and protection jurisprudence in the long run is what one needs to watch out for.

Source: http://www.business2community.com/big-data/minority-report-confronting-privacy-issues-big-data-gathering-0970881#S50GIf5JMYS7v4tj.99.

11. Conclusion

While in Big Data technology costs are low due to the presence of open source tools, the major cost is the human resources required for planning, cultural alignment, process definitions and deployment.

Studies on Big Data analytics have indicated that about 20% of companies using data analytics are able to outperform their competitors. Close to 80% of internet marketing companies are in the process of implementing Big Data initiatives.

Deployment of Big Data strategies can be challenging, but the really encouraging news about Big Data projects is that they generally deliver big paybacks. A recent Nucleus Research report titled "The Big Returns from Big Data" found that Big

Data projects which connected internal data sets with social media earned, on average, 241 percent ROI.

Big Data is yet to cross the chasm to mainstream adoption, but is clearly delivering success for early adopters and is now at an inflexion point. For most businesses, Big Data methods are as unique as their corporate cultures and business processes and the technologies carry a big hype in today's business environment.

Bibliorgraphy

1. Vesset, D., Morris, H.D., Little, G., Borovick, L., Feldman, S., Eastwood, M., Woo, B., Villars, R.L., Bozman, J.S., Olofson, C.W., Conway, S., &Yezhkova, N.
2. IDC market analysis: Worldwide Big Data technology and services, IDC #233485, March 2012.
3. Duhigg, C. How companies learn your secrets, The New York Times, 2/16/2012.
4. Hoffman, D.L., & Fodor, M. (2010). Can you measure the ROI of your social media marketing?, MIT Sloan Management Review, 52(1), 41-49.
5. Schaefer, M.W. (2012). Return on influence, McGraw-Hill, pp 5-6.
6. IDC Customer Spotlight: Whirlpool corporation's digital detectives: Attensity provides the lens, March 2011.
7. Mckinsey quarterly May 2013-Big Data what is your plan?
8. Data analytics: integration and privacy by Steven E. Whang, June 2012
9. Big Data = Big Opportunity = Big Problems? By BrettKeirstead Vice President Sales Knowledge Marketing brett.keirstead@teamkm.com
10. Effects of Big Data analytics on Organizations' value creation by Niels Mouthaan,,Msc, University of Amsterdam, Faculty of science and Faculty of economics and business
11. META Group. "3D Data Management: Controlling Data Volume, Velocity, and Variety." February 2001.:Thomas H. Davenport and Jill Dyche, "Big Data in Big Companies," May 2013.

Web pages referred

https://en.bitcoin.it/wiki/Scalability,

http://expandedramblings.com/index.php,

http://www.complex.com/tech/2012/10/twitter,

http://www.crmsearch.com/big-data-use-cases.php,

http://www.crmsearch.com/big-data-implementation.php

***