BIG DATA EXPLOSION: INSIGHT FOR NEW AGE MANAGERS

Abhinav Bansal, Student of IV year, Dept. of Civil Engineering, JUIT
Dr. Anupriya Kaur, Assistant Professor, Dept. of Humanities and Social Science, JUIT
Ashish Aggarwal, Student of III year, Dept. of Computer Science and Engineering, JUIT

ABSTRACT: There is an evident shift in decision – making processes and we have evolved from the times of making crude estimates and judgments to painstakingly (time consuming) constructed models simulating reality to the data-driven real time decision-making in the form of "Big Data". We are awash in a flood of data today. In a broad range of application areas, data is being collected at unprecedented scale. Such Big Data analysis now drives nearly every aspect of our modern society, including mobile services, retail, manufacturing, financial services, life sciences, and physical sciences. Thus, it is an area which interests academicians, researchers and practitioners alike. This paper is aimed to present a synoptic analysis on the potential and outreach of the Big Data technologies. Further, it focuses on the trends, opportunities and challenges associated with the same. The central theme of the paper is to provide the new age practitioners insights on the promising aspects of the Big Data technologies which can equip the organization with strategic advantages in operational efficiencies and augmenting their market offerings. Big Data has the potential to revolutionize not just research, but also education. It is widely believed that the use of information technology can reduce the cost of healthcare. Heterogeneity, scale, timeliness, complexity, and privacy problems with Big Data impede progress at all phases of the pipeline that can create value from data.

INTRODUCTION

Increasing amounts of data are streaming into contemporary organizations as a result of the rapidly growing quantity of data being generated not only by the organizations themselves but also in the organizations' business environments by both their stakeholders and other entities operating there. Thus, it is in this context that such expressions as "a data-centric world" have become more and more common [1].

Big Data is emerging from the realms of science projects at Web companies to help companies like telecommunication giants understand exactly which customers are unhappy with service and what processes caused the dissatisfaction, and predict which customers are going to change carriers. To obtain this information, billions of loosely-structured bytes of data in different locations needs to be processed until the needle in the haystack is found. The analysis enables executive management to fix faulty processes or people and maybe be able to reach out to retain the at-risk customers. The real business impact is that big data technologies can do this in weeks or months, four-or-more-times faster than traditional data warehousing approaches. Big data is data that is too large to process using traditional methods. It originated with Web search companies who had the problem of querying very large distributed aggregations of loosely-structured data. [2]

Big Data: Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges. [3]

Big data has four major characteristics 4 V's:

1. **Volume** : Scale of data

Enterprises are swimming in data, banks, for example collect data in terabytes (the U.S. Library of Congress' total book stack measures 15TB – and petabytes (Google processes more than 1PB every hour). Managing huge volumes of data is a major challenge for financial services firms, for example. Data sharing across Wall Street enterprises is still a big issue, as each business unit frequently prefers calculating from its own set of data. With these companies collecting even more unstructured data, advanced enterprises have developed tools that can analyze news, via video, audio and Twitter, for example, in real-time to help make trading decisions. New regulations focused on transparency and risk management to be put in place in 2012 are driving greater urgency among capital markets firms to manage big data. "Big data in financial services in 2012 – or 2015, or 2020 for that matter – is going to be an important topic," said James Austin, CEO of Vertex Analytics. "There aren't going to be any firms that want less data in the future; they are all going to want more data. And regulations are going to play a large part in big data," he told Wall Street & Technology magazine in November, 2011. Bank of America Merrill Lynch, for example, is using Hadoop technology to manage petabytes of data for advanced analytics and new regulatory requirements.

2. **Velocity** : Analysis of streaming data.

Global banks handle trillions of messages in a single day's trading, mostly processed by computers.

3. **Veracity** : Uncertainty of data.

Semantics, or the variability of meaning in language.

4. **Variety** : Different forms of data

The IT industry has dealt with big data for decades as structured data in static and disciplined databases and spreadsheets. What's new are tools to effectively capture, visualize and analyze unstructured data that is messy, moving, ubiquitous, streaming in text, audio, video, clicks, PDF files, email, blogs, tweets, sensors and the rest. About 80 percent of a corporation's data is unstructured including office productivity documents, e-mail, Web content, in addition to social media. Email and messaging systems create unstructured data more than anything else. While two of five respondents to a Unisphere survey say upper management is barely aware of the challenges of unstructured data, IT professionals are seriously concerned about the volumes they're getting. At least 57 percent of respondents report that unstructured data is very important, and about 18 percent consider it a core of their business.

Big data could be a mix of digital video, text (increasingly coming from comments on social media sites such as Twitter, Facebook, LinkedIn etc.), digitized audio and other types. To analyze this data, the technology needs to process it in some manner. [4]

Creating value out of Big Data

According to the McKinsey Global Institute[6], five key ways in which Big Data creates value for organizations can be distinguished:

1. Creating transparency by integrating data and making it more easily accessible to all relevant stakeholders.
2. Enabling experimentation to discover needs, expose variability, and improve performance.
3. Segmenting populations in order to customize actions,
4. Replacing or supporting human decision making with automated algorithms.
5. Innovating new business models, products and services.

Success factor of Big Data

Marchand and Peppard [7] have identified five important guidelines for the success of a Big Data project. They include

1. Placing people at the heart of the Big Data initiative.
2. Emphasizing information utilization as the way to unlock value from information technology.
3. Equipping IT project teams with cognitive and behavioral scientists.
4. Focusing on learning.
5. Worrying more about solving business problems than about deploying technology.

According to Barth et al., [8] organizations that benefit from Big Data base their activities on following three fundamental issues:

1. Paying attention to data flow as opposed to stocks.
2. Relying on data scientists and product and process developers rather than data analysts.
3. Moving analytics away from the IT function, into core business, with operational and production functions.

Barton and Court, [9] on the other hand, came to the conclusion that full exploitation of data and analytics requires three capabilities:

1. Choosing the right data. In this context two aspects are important: creative sourcing of internal and external data and upgrading IT architecture and infrastructure for easy data merging.
2. Emphasizing information utilization as the way to unlock value from information technology. In this context two aspects are important: focusing on the biggest drivers of performance and building models that balance complexity with ease of use.
3. Equipping IT project teams with cognitive and behavioral scientists. In this context two aspects are important: creating simple, understandable tools for people on the front lines and updating business processes and developing capabilities to enable tools utilization.

Challenges in Creating Business Value Out of Big Data

1. **Sharing Information**: Getting business units to share information across organizational silos.
2. **Three V's**: Being able to handle the large volume, velocity and variety of Big Data.
3. **Identifying data**: Determining what data (both structured and unstructured, and internal and external) to use for different business decisions.
4. **Trust Building**: Building high levels of trust between the data scientists who present insights on Big Data and the functional managers.
5. **Human Resource**: Finding and hiring data scientists who can manage large amounts of structured and unstructured data and create insights. Reskilling the IT function to be able to use the new tools and technologies of Big Data. Getting functional managers to make decisions based on Big Data.
6. **Financial Support**: Getting top management in the company to approve investments in Big Data and its related investments (e.g., training, etc.).
7. **Analysis**: Putting our analysis of Big Data in a presentable form for making decisions (e.g., visualization/visual models) and finding the optimal way to organize Big Data activities in the company. Understanding where in the company we should focus our Big Data investments.
8. **Application**: Determining what to do with the insights that are created from Big Data. Determining which Big Data technologies to use.
9. **Security Issues**: Keeping the data in Big Data initiatives secure from external parties and internal parties.[4]

Application of Big Data

Amazon's retail price competition:

Amazon burst into the headlines in December, 2011 with a promotion offering iPhone and Android users $5 off for sharing in-store prices while shopping for toys, electronics,

sports, music, and DVDs, with their mobile phone application. The twofold aim was to increase usage of its Amazon's Price Check App bar-code scanning application while also collecting comparative intelligence on store prices. Using the Amazon app. shoppers scan a bar code, take a picture of the item or conduct a text search to find the lowest price.

Fed BP Disaster response:
Early in 2010's Deepwater Horizon disaster, the oil rate flow was a major question. BP and independent groups offered varying estimates hindering efforts to coordinate the scale and scope of the U.S. Government's response. With close enough not good enough, the National Institute of Science and Technology (NIST), worked to make sense of the disparate estimates. NIST used the open source R statistics language to analyze and reconcile the estimates and produced actionable intelligence on which to base the ultimate response.

Analyzing social media streams:
One company measures influence across the social web by storing, processing, and analyzing real time social media data streams. The company's platform analyzes signals as they travel through the social Web and performs natural language processing, machine learning, and other analysis to measure topical and broad based influence.

Cross-selling:
One company increases an average order size by recommending complementary products based on predictive analytics for cross selling.

Email:
An insurance company implements big data analytics to retain emails and other electronic documents in anticipation of future litigation and investigations. In the case of a legal hold placed on the company's electronic information, or an internal investigation the archived data is searched and delivered within the system. The capacity to speedily access vast quantities of previously inaccessible electronic information due to its unstructured format accelerates the legal process, aids investigative analysis, and improves the odds of mounting a successful case.

Fraud:
Governments and credit card companies are detecting fraud (including claims and tax) in online systems in real-time based on behavioral patterns.

Human behavior:
Companies are able to quantify aspects of human behavior never before accessible. Social networks, news stream, and the smart grid, are way of measuring "conversation", "interest", and "activity". What's more, with data machine-learning algorithms and big data tools to reveal insight, it used only the structure, not the content, of the data. One company has collected 3 billion tweets from 60+ million users tokenized into 16 billion+ usages of 65 million terms – more than a terabyte of data. It can identify whom to follow to understand how events and news stories resonate, and even to find dates.

Keyword campaigns:
Retailers use big data analytics to drive traffic from search engines to their Web sites. The software collects information about which keywords work best to turn Internet surfers into shoppers. They also create language models so that their sites can return product results if the shopper enters a related word. For example, if a shopper searches for "dining table" on a retail site it will return results for dining room furniture. The retailer may be able to return in future to offer new styles of dining room furniture more likely to appeal based on that particular shoppers' tweets and Facebook updates.

Location tracking:
A company collects information for its mapping service it sells to large businesses. The company can tap into data from probes and mobile devices around the world to collect traffic data. To figure out information about a particular street, the company formerly had staff poring through hundreds of terabytes of data.

Market and customer segmentation:
An arms race is occurring in the retail sector. If retailers can understand consumer behavior and collect behavioral data to better guide product decision-making, then every cent they can eke out is increasing their margins, and allowing them to invest more.

Monitoring mobile phone usage:
Usage information such as average screen brightness, signal strength, wifi connectivity, power cycles, and more is collected daily and sent to a cluster. Data is sent when the phone is connected to a power source to avoid using battery to send data upstream. Once the company's monitor has been running for a few weeks on a phone, it starts analyzing the usage data and makes recommendations in the form of push notifications. A notification might suggest turning on auto screen brightness, or start using wifi when the signal is low.

Policing power tools:
Using big data analytics makes for more dynamic law enforcement. In North Carolina police are using a Web based application that offers courts and law enforcement agencies access to integrated criminal data. A view of an offender is seen through a single application, allowing for positive identification through a photograph as well as an "offender watch" capability to alert law enforcement professionals of an offender's change in status. The servers integrate data within a state's criminal justice applications.

Managing risk:
The North Carolina Department of State Treasurer is turning to analytics software to manage the state's pension risk to better protect the pensions of more than 850,000 residents. The risk management technology, customized for state

pensions, will help offices in the department's Investment Management Division better assess the risks associated with new and current investments in its $72.8 billion portfolio.

Risk modeling:
A large bank combined separate data warehouses from multiple departments into a single global repository for analysis. The bank used the cluster to create a new, more accurate score of the risk into its client portfolios. This enables the bank to better manage its exposure, and to offer its clients better products and advice.

Tracking ads:
A critical premise for success of online advertising networks is to successfully collect, organize, analyze and use large volumes of data to place an ad appropriate to a Web page user. Given the nature of their online orientation and dynamics, it is critical that these processes be automated to the largest extent possible. Specifically, the success of advertising technology and its impact on revenue are directly proportional to its capability to use large amounts of data in order to compute proper impression value given the unique circumstances of ad serving events such as the characteristics of the impression, the ad, and the user as well as the content and context. As a general rule, more data results in more accurate predictions.

Customer churn:
A large telecommunications company provided analyzed call logs and complex data from multiple sources. Using predictive analytics across this data it was able to predict the likelihood of any individual customer cancelling (churn). The information also helped the company create more sensitive relationships with customers to reduce churn.

Activities that can be enriched by Big Data

1. **Sales**:
   a. Identifying customers with the most value/potential value.
   b. Identifying cross-selling opportunities.
   c. Determining optimal sales approaches, optimal sales offers & optimal sales messages (for example, from sales wins and losses)
   d. Pricing.

2. **Research & Developement**:
   a. Monitoring product quality
   b. Identifying customer needs for new products and enhancements to existing products
   c. Identifying spikes in logistics costs and where and why they are occurring.
   d. Getting continuous customer feedback on products already in the market.
   e. Testing new product designs.

3. **Logistics**:
   a. Monitoring product shipments.
   b. Determining locations of inventory shrinkage.
   c. Boosting energy efficiency.
   d. Identifying supply chain bottlenecks to speed flow of goods, materials.
   e. Determining appropriate inventory levels.

4. **Customer service**:
   a. Identifying customers at risk of dropping our products/services.
   b. Analysing customer behaviour on the website to see which pages are most and least useful.
   c. Identifying patterns in customer complaints (both internal in the call centers, and external) **and** customer inquiries (for example, to better allocate service personnel to peak times)

5. **Marketing**:
   a. Determining marketing campaign & channel effectiveness.
   b. Determining customer value
   c. Tailoring marketing campaigns and promotional offers.

6. **Human Resource**:
   a. Improving employee retention by determining who is most likely to leave.[4]

Conclusion
The world has seen the advent of big data and its computing as a strategy which will set the future benchmarks for not only enterprises and organizations but also for each and every field which might get influenced by certain trends and approximations. This itself makes this an interesting subject to explore and research. With the increasing window of its applications from security to business management and marketing, big data analysis has a wide range of applications which is steadily growing day by day. But this whole concept of processing data efficiently and uniquely has some fine challenges of data being consistent, secure, veracity and many more. Its mathematical and statistical computation also poses some serious challenges of varied applications on the data. These challenges are emerging as potential research fields and thus are giving diverse edges to this whole new domain. In near future big data analytics will play a significant role in decision making as it gets wide acceptance in various frays of its application, and will continue to evolve as it gets more big and complex. As more and more data gets collected and stored, its analytics and computation will definitely be proved to some significant use in judging out the future paradigms.

REFERENCES

1. Kemp Little LLP, "Big Data – Legal Rights and Obligations",http://www.kemplittle.com/Publications/WhitePapers/Big%20Data%20%20Legal%20Rights%20and%20Obligations%202013.pdf, January, 2013.

2. Foyer, David & Vellante, David (2014), Enterprise Big-Data, http://wikibon.org/wiki/v/Enterprise_Big-data, April 6, 2014.
3. The OED, Big Data, and Crowdsourcing | What's The Big Data? http://whatsthebigdata.com/, August 17, 2013.
4. Emerging big returns on big data (2013), TCS Global Trend Survey 2013, Tata consultancy service (TCS).
5. AMD white paper: Big data- It's not just for google any more. The software and compelling economics of big data computing.
6. McKinsey Global Institute, "Big data: The next frontier for innovation, competition, and productivity", http://www.mckinsey.com/mgi/publications/big_data/pdfs/MGI_big_data_full_report.pdf, May 2011.
7. D. Marchand, J. Peppard, "Why IT fumbles analytics", Harvard Business Review, pp. 104-113, January-February 2013.
8. P. Barth, R. Bean, T. Davenport, "How Big Data is Different", Sloan Management Review, Fall 2012.
9. D. Barton, D. Court, "Making advanced analytics work for you", Harvard Business Review, pp. 78-83, October 2012.