



## Big Data Analytics Challenges and Opportunities Ahead – Beyond the Hype

**Abirami N**

Assistant Professor  
Department of Information Technology  
Karpagam Academy of Higher Education,  
Coimbatore, India

**K. Veerasamy**

Assistant Professor,  
Department of Computer Applications  
Karpagam Academy of Higher Education,  
Coimbatore, India

### ABSTRACT

The term “Big data” itself defines the voluminous data and analytics of the big data that requires smart collection, handling and utilization. With the evolution of technology this might look simple but by virtue of its existence and elegance in its availability it poses a big challenge. This paper aims to bring together the face of the past, present and future aspects big data analytics in the light of its various challenges.

**Keyword:** big data, big data process, V's , advantages, disadvantages, do's, don'ts, challenges.

### INTRODUCTION

Huge volume of data, both structured and unstructured, have been generated over the decade with the evolution of high-end technologies like smart phones, which loaded with mega pixel cameras. These data are high in capacity and cannot be handled with traditional tools for the information processing. The large variable data with high volume, velocity and complexity are called Big Data[2]. These data require technology that can manage them for storage and information analysis for future utilization in decision making processes.

Volume, Variety, Velocity, Veracity, Variability, Value, Venue, Vocabulary, Vagueness and Validity are the V's that have emerged as a common framework to describe big data.

Big data availability had to be put into maximum information retrieval process. The technique that is used to analyze and acquire such information retrieval called Knowledge from big data is called Big data Analytics. Many enterprises come across facts that they were unaware of in the past. With awareness spreading everywhere, many organizations are now implementing this Big data analytics to marching forward and stay connected and be ahead in the evolving competitive world.

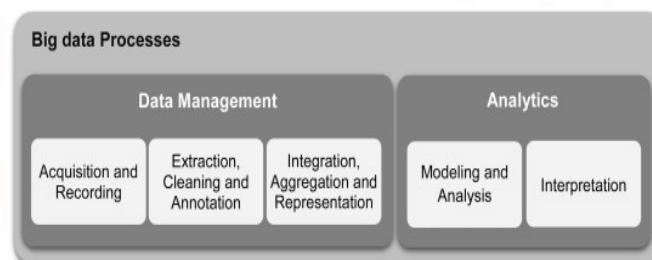


Fig. 1: Steps for extracting insights from big data

Imagine the data that is available today. For example, large and voluminous satellite images captured by orbiting satellites transmit them to earth stations every few seconds. Also mass transfer of data in the social media like twitter, facebook, whatsapp, email and from all internet of things to small scale purchase of grocery items by people everyday has become indispensable. The business efficiency of any organization, big or small, depends on super fast transfer of data.

Data available can be categorized broadly as structured and unstructured form. Textual data available from social networks, emails, blogs, online forums, survey responses, corporate documents, news and call centre logs are examples of textual data held by organizations which can be in both structured and unstructured forms. Unstructured audio data are those in the form of speech through recorded audios or live calls in call centres, video data available through camera recorders and still images available through capture of objects.

Data required for problem solving need to be the right data rather than big data. Unstructured data constitute 95% of big data. Executives and others differ in the way they understand, define and focus on what big data is and what it does.

To understand this and put it into best practise, one needs to first understand the do's, don'ts and challenges that surround the big data usage. After understanding these basics, implementing and practising them become easier.

### **Effects of big data in our daily lives.**

- Sports predictions- predicting outcomes of sport events
- Voting predictions – predicting outcomes of elections
- Smart phones - direction enquiry through voice and text
- Personalized advertising and purchasing recommendations – purchasing advertisements on the web
- Improved traffic flow – streamlining of traffic flow to prevent accidents
- Epidemic detection and prevention – outbreak of potentially epidemic viruses such as flu.<sup>[12]</sup>

### **Advantages**

- Enabling of new directions towards the area of scientific research that was constrained by the volume of data.
- Small data samples that create malicious problems can be avoided by fitting them in the right models using test data for quality validation and better handling of noisy training data

### **Disadvantages**

- Lack of standards and limited or missing theoretical bases for data models and query languages that would help to avoid vendor lock-in.
- Requirement of fast and approximated algorithms where precise solutions are not feasible to those problems that might degrade the quality of deriving a solution. Lack of tradeoffs among desired scalability, availability, performance and security on data poses a big problem.

### **Do's**

- Have a strong leader at the top for decision making and implementations to drive initiatives taken by using analytics-driven big data<sup>[1]</sup>.
- Have multiple channels to build big data capabilities
- Build an organization centric “big data lab” that offer big data complete
- Build the right data model with appropriate tools, security and privacy.
- Follow the iterative approach towards implementation.
- Establish good governance framework, procedures and policies for managing data assets.
- Develop competencies

### **Don'ts**

- Do not store pool of data that goes waste and is irretrievable.
- Do not invest more on security unless essential thinking is mandatory<sup>[1]</sup>.
- Do not use inconsistent data for analysis and decision making.
- Do not use non-analysed and unsorted data, before it is used for useful information.

### **Challenges**

- Decision makers need to be able to gain valuable insights from such varied and rapidly changing data<sup>[2]</sup>.
- Big Data are data sets whose size is beyond the ability of commonly used software tools and storage systems to capture, store, manage, as well as process within a tolerable elapsed time
- The size of Big data are constantly increasing
- Capture, transfer, store, clean, analyze, filter, search for, share, secure, retrieve and visualize data.

- Follow a problem-driven approach to those data that are noisy, biased, imprecise, incomplete, subjective, redundant and conflicting rather than a data-driven one.
- Big data are worthless in a vacuum. Their potential value is unlocked only when leveraged towards decision making.
- The key characteristic of the modern social media analytics is its data-centric nature and big data technologies to be adopted to address the data processing challenges.
- Highly heterogeneous data that represent information from different sources and different available population.
- People have not understood the term “big data” clearly
- Spending on Internet of things will grow by \$2.5 million per minute that can connect 30 million objects<sup>[8]</sup>.

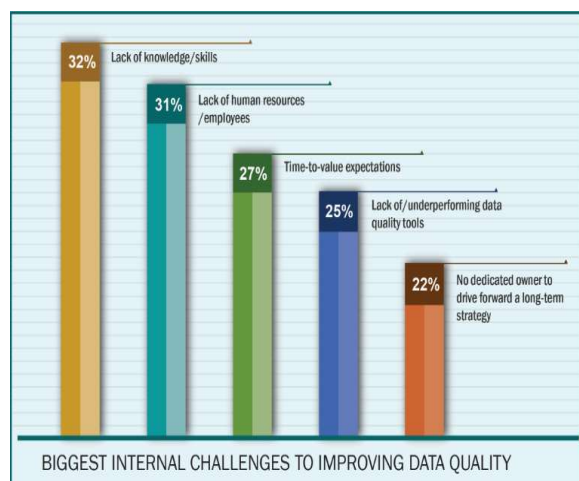


Fig. 2 Biggest Internal Challenges to Improving Data Quality

### Future:

Well defined criteria, timeline with a roadmap are essential for success. Storage capacities will increase, allowing even bigger data sets to be captured With nearly 60% of industry persons realizing the fact that big data will disrupt their growth, lack of strong data management and governance are major challenges ahead. Global organizational spending on Big Data exceeded \$31 billion in 2013, and is predicted to reach \$114 billion in 2018. An analysis by experts predicts that around 40 zettabytes of data will be in existence by 2020. which is nearly 140% in comparison to data that was available in 2013<sup>[7]</sup>.

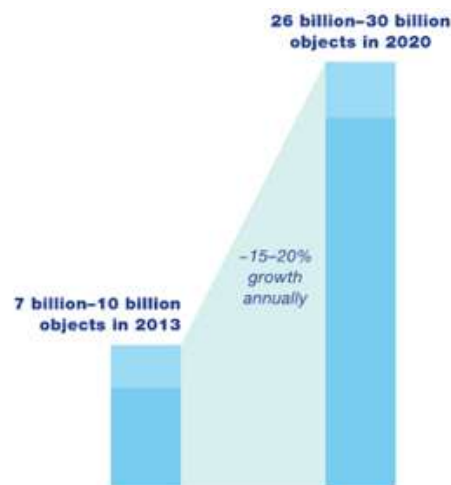


Fig. 3 Increase in number of objects that can be connected by 2020 through IoT

### **References**

- 1) Stephanie Overby, August 21, 2012, Six Do's and Don'ts for Outsourcing Big Data Analytics.
- 2) Karthik Kambatla, GiorgosKollias,VipinKumar, Ananth Grama, July 2014, Trends in big data analytics, Journal of Parallel and Distributed Computing, 74(7), 2561-2573
- 3) Esther Shein, January 23, 2017, Insights and resources to help IT pros shape the future of business.
- 4) Bhavna Singh, October 24, 2015, Top 10 Amazing Facts To Know About Big Data.
- 5) Daniel Price, March 17, 2015, Surprising facts and stats about the big data industry, Cloud tweaks.
- 6) APRE, January 23,2017, Intro session big data content pomeriggio.
- 7) Aatash Shah, December 23, 2015, 7 big data trends 2016, Edvancer
- 8) Stratebi, August 29, 2013, Big data analytics 2014, slideshare.net/zanorte/category/business
- 9) Saint John Walker, Jan 2015, Big Data: A Revolution That Will Transform How We Live, Work and Think, International Journal of Advertising The Review of Marketing Communications, 33(1):181-183.