ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 613 - 615 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

Big Data Challenges

[1] **Nrusimham Ammu,** [2] **Mohd Irfanuddin**
M.Tech Students, LIET,Hyderabad

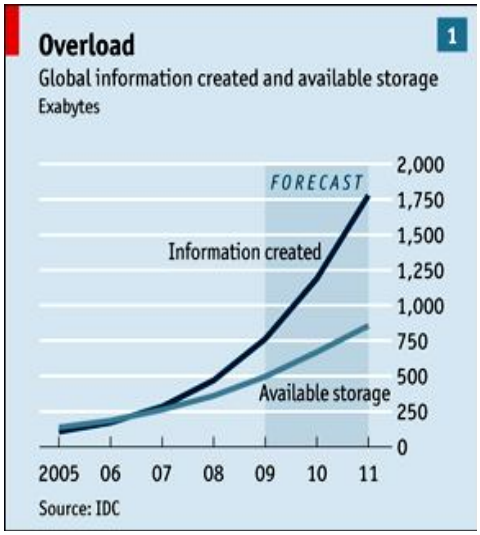
ABSTRACT

Big Data, an umbrella term for the explosion in the quantity and diversity of high frequency digital data. The Big Data may be logs, mobile-banking transactions, online user-generated content such as blog posts and Tweets, online searches, satellite images, etc.- into actionable information requires using computational techniques to unveil trends and patterns within and between these extremely large socioeconomic datasets. These data hold the potential - to allow decision makers to track development progress, improve social protection, and understand where existing policies and programmes require adjustment. This paper presents the novel challenges and opportunities associated with Big Data necessitate rethinking many aspects of these data management platforms, while retaining other desirable aspects.

Keywords: Logs, GPS, power budget, Visualization.

INTRODUCTION

Big Data is a popular phrase used to describe a massive volume of both structured and unstructured data that is so large that it's difficult to process with traditional database and software techniques. The characteristics which broadly distinguish Big Data are the "3 V's": more volume, more variety and higher rates of velocity. In earlier a relatively small volume of analog data was produced and made available through a limited number of channels, today a massive amount of data is regularly being generated and flowing from various sources, through different channels, every minute in today's Digital Age. In fact , today, we are afloat in Data Ocean. In a broad range of application areas, data is being collected at unprecedented scale. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos posted online, transaction records of online purchases, and from cell phone GPS signals. It is the speed and frequency with which data is emitted and transmitted on the one hand, and the rise in the number and variety of sources from which it emanates on the other hand, that jointly constitute the data surge. The amount of available digital data at the global level grew from 150 exabytes(1eb= billion gb) in 2005 to 1200 exabytes in 2010. It is projected to increase by 40% annually in the next few years, which is about 40 times the much-debated growth of the world's population. This rate of growth means that the stock of digital data is expected to increase 44 times between 2007 and 2020, doubling every 20 months.



Source: "The Leaky Corporation." *The Economist*.
http://www.economist.com/node/18226961

Value of Big Data:

a) In the biological sciences, there is now a well-established tradition of depositing scientific data into a public repository, and also of

ISSN 2278-3091

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 613 - 615 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

creating public databases for use by other scientists.

b) It is widely believed that the use of information technology can reduce the cost of healthcare while improving its quality, by making care more preventive and personalized and basing it on more extensive (home-based) continuous monitoring. McKinsey estimates [McK2011][1] a savings of 300 billion dollars every year in the US alone.

c) The Big Data has potentially revolutionize the urban Planning[2], intelligent transportation, (through analysis and visualization of live and detailed road network data), environmental modeling (collecting data through sensor networks in all places) , energy saving (through unveiling patterns of use), smart materials (through the new materials genome) initiative computational social sciences

CHALLENGES

While the potential benefits of Big Data are real and significant, many technical challenges that must be addressed to fully realize this potential.

i) **Heterogeneity and Incompleteness**

The computer systems work most efficiently if they can store multiple items that are all identical in size and structure. Efficient representation, access, and analysis of semi-structured data require further work.

One big challenge is to automatically generate the right metadata to describe what data is recorded and how it is recorded and measured.

A problem with current Big Data analysis is the lack of coordination between database systems, which host the data and provide SQL querying, with analytics packages that perform various forms of non-SQL processing, such as data mining and statistical analyses.

Scale

For many decades managing large and rapidly increasing volumes of data has been a challenging issue . In the past, this challenge was alleviated by processors getting faster, following Moore's law, to provide us with the resources needed to cope with increasing volumes of data. But, there is a fundamental shift happening now: data volume is scaling faster than compute resources, and CPU speeds are static. Now, due to power constraints, clock speeds have largely stalled and processors are being built with increasing numbers of cores. with predictions of "dark silicon"(Dark Silicon refers to the exponentially increasing number of a chip's transistors that must remain passive, or "dark", in order to stay within a chip's power budget), namely that power consideration will likely in the future prohibit us from using all of the hardware in the system continuously. Data processing systems will likely have to actively manage the power consumption of the processor. These unprecedented changes require us to rethink how we design, build and operate data processing components.

Timeliness

Longer time needs to analyze the larger data set. However, there are many situations in which the result of the analysis is required immediately. Given a large data set, it is often necessary to find elements in it that meet a specified criterion. It is impractical to Scanning the entire data set to find suitable elements . Rather, index structures are created in advance to permit finding qualifying elements quickly. In doing so, each index structure is designed to support only some classes of criteria. When new analyses desired using Big Data, there are new types of criteria specified, and a need to devise new index structures to support such criteria. New index structures are required to support such queries. Designing such structures becomes particularly challenging when the data volume is growing rapidly and the queries have tight response time limits.

We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available, there is the potential for making faster advances in many scientific

International Journal of Advanced Trends in Computer Science and Engineering, Vol.2 , No.1, Pages : 613 - 615 (2013)
Special Issue of ICACSE 2013 - Held on 7-8 January, 2013 in Lords Institute of Engineering and Technology, Hyderabad

disciplines. However, many technical challenges described in this paper must be addressed before this potential can be realized fully. The challenges include not just the issues of scale, but also heterogeneity, lack of structure, error-handling, privacy, timeliness, provenance, and visualization, at all stages of the analysis pipeline from data acquisition to result interpretation. These technical challenges are common across a large variety of application domains, and therefore not cost-effective to address in the context of one domain alone. Furthermore, these challenges will require transformative solutions, We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data.

CONCLUSION

Big Data is an important phenomena like nanotechnology and quantum computing in the present era. Big Data is the umbrella term for the Data , though for unstructured, semi structured and structured. Through better analysis of the large volumes [3] of data that are becoming available, that are many areas that are unexplored. However, many technical challenges as described in this paper must be addressed, before the fruits of vast realm of big data available to us.

REFERENCES

[1].[McK2011] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011.

[2]"Big Data, Big Impact: New Possibilities for International Development." *World Economic Forum*

[3] Big data: The next frontier for innovation, competition, and productivity. James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. McKinsey Global Institute. May 2011.