

# Databig data and future of data-driven innovation



**ASSIGN  
BUSTER**

DataBig Data and Future of Data-Driven Innovation A. A. C. Sandaruwan

Faculty of Information Technology University of Moratuwa

chanakasan@gmail. com The section 2 of this paper discuss about real world

examples of big data application areas. The section 3 introduces the

conceptual aspects of Big Data. The section 4 discuss about future and

innovations through big data. Abstract: The promise of data-driven decision-

making is now being recognized broadly, and there is growing enthusiasm

for the notion of ``Big Data.'' Heterogeneity, scale, timeliness, complexity,

and privacy problems with Big Data impede progress at all phases of the

pipeline that can create value from data. Much data today is not natively in

structured format; for example, tweets and blogs are weakly structured

pieces of text, while images and video are structured for storage and display,

but not for semantic content and search: transforming such content into a

structured format for later analysis is a major challenge. The value of data

explodes when it can be linked with other data, thus data integration is a

major creator of value. 2. Big Data in the Real World Big Data talks about this

increasing amounts of data available for companies that can be used to

capture value. In simplest terms, the phrase refers to the tools, processes

and procedures allowing an organization to create, manipulate, and manage

very large amounts of data. It does not define how much is big; it depends

on the context, as what one company considers big could be relatively small

for another company. So this refers to data that is large enough that our

traditional tools will struggle to handle not whether it's terabytes or

petabytes of data. Since most data is directly generated in digital format

today, we have the opportunity and the challenge both to influence the

creation to facilitate later linkage and to automatically link previously

<https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>

created data. Data analysis, organization, retrieval, and modeling are other foundational challenges. Data analysis is a clear bottleneck in many applications, both due to lack of scalability of the underlying algorithms and due to the complexity of the data that needs to be analyzed. A major investment in Big Data, properly directed, can result not only in major scientific advances, but also lay the foundation for the next generation of advances in science, medicine, and business. 2. 1. Opportunities 1.

Introduction Scientific research has been revolutionized by Big Data.

Astronomy is being transformed from one where taking pictures of the sky was a large part of an astronomer's job to one where the pictures are all in a database already and the astronomer's task is to find interesting objects and phenomena in the database [1, 2]. The widespread use of the internet has unveiled endless possibilities for many different aspects of the society. The increasingly large amounts of data available through the internet has made it challenging for companies to meet the ever evolving needs of today's society. While the availability of large amounts of data has created limitless opportunities for businesses and IT companies; their traditional methods of handling data are insufficient to explore these new possibilities. Big Data has the potential to revolutionize not just research, but also education. Imagine a world in which we have access to a huge database where we collect every detailed measure of every student's academic performance. This data could be used to design the most effective approaches to education, starting from reading, writing, and math, to advanced, college-level, courses. We are far from having access to such data, but there are powerful trends in this direction. In particular, there is a strong trend for massive Web deployment of educational activities, and this will generate an increasingly large amount

<https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>

of detailed data about students' performance. The notion of Big Data has emerged to meet these demands and requirements. With Big Data decisions that were previously made based on guesswork, or on models of reality created by complicated methods can now be based on the data itself. The requirement for data-driven decision-making is becoming crucial factor for every aspect of modern society, including mobile services, financial services, life sciences and physical sciences. It is widely believed that the use of information technology can reduce the cost of healthcare while 1 improving its quality, by making care more preventive and personalized and basing it on more extensive monitoring. The phases of big data analysis include, - - Similarly, the SQL standard and the relational data model provide a uniform, powerful language to express many query needs and, in principle, allows customers to choose between vendors, increasing competition. The challenge ahead of us is to combine these healthy features of prior systems as we devise novel solutions to the many new challenges of Big Data. 3. 1. Extracting Big Data Frequently, the information collected will not be in a format ready for analysis. For example, consider the collection of electronic health records in a hospital, comprising transcribed dictations from several physicians, structured data from sensors and measurements (possibly with some associated uncertainty), and image data such as x-rays. We cannot leave the data in this form and still effectively analyze it. 3. 2. Big Data Integration Large portions of data found in datasets gathered by companies happen to be unstructured data. Data analysis is considerably more challenging than simply locating, identifying, understanding, and citing data. For effective large-scale analysis all of this has to happen in a completely automated manner. This requires differences in data structure and semantics

<https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>

to be expressed in forms that are computer understandable, and then “robotically” resolvable. There is a strong body of work in data integration that can provide some of the answers. However, considerable additional work is required to achieve automated error-free difference resolution. When processing these data the following criteria have to be considered.

Heterogeneity Scale - Timeliness - Privacy - Interpretation Fortunately, existing computational techniques can be applied, either as is or with some extensions, to at least some aspects of the Big Data problem. For example, relational databases rely on the notion of logical data independence: users can think about what they want to compute, while the system (with skilled engineers designing those systems) determines how to compute it efficiently.

3. Conceptual aspects behind Big Data - Analysis - The three design choices listed have successively less structure and, conversely, successively greater variety. Greater structure is likely to be required by many (traditional) data analysis systems. However, the less structured design is likely to be more effective for many purposes — for example questions relating to disease progression over time will require an expensive join operation with the first two designs, but can be avoided with the latter. However, 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. -

Integration - When humans consume information, a great deal of heterogeneity is comfortably tolerated. In fact, the nuance and richness of natural language can provide valuable depth. However, machine analysis algorithms expect homogeneous data, and cannot understand nuance. In consequence, data must be carefully structured as a first step in (or prior to)

<https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>

data analysis. Consider, for example, a patient who has multiple medical procedures at a hospital. We could create one record per medical procedure or laboratory test, one record for the entire hospital stay, or one record for all lifetime hospital interactions of this patient. With anything other than the first design, the number of medical procedures and lab tests per record would be different for each patient.

## 2. 2. Challenges Acquisition

### Human Collaboration

The analysis of Big Data involves multiple distinct phases [6]. Many people unfortunately focus just on the analyzing phase while that phase is crucial, it is of little use without the other phases of the data analysis pipeline.

### 3. 3. Big Data Analysis Methods for querying and mining Big Data [7]

are fundamentally different from traditional statistical analysis on small samples. 2 Big Data is often noisy, dynamic, heterogeneous, interrelated and untrustworthy. Nevertheless, even noisy Big Data could be more valuable than tiny samples because general statistics obtained from frequent patterns and correlation analysis usually overpower individual fluctuations and often disclose more reliable hidden patterns and knowledge. Further, interconnected data “ the next frontier for innovation, competition and productivity” [3, 4]. We can answer questions with big data that were beyond reach in the past. We can extract insight and knowledge, identify trends and use the data to improve productivity, gain competitive advantage and create substantial value for the world economy. The challenges with big data are limited compared to the potential benefits, which are limited only by our creativity and ability to make connections among the trillions of bytes of data we have access to. Big Data forms large heterogeneous information networks, with which information redundancy can be explored to compensate for missing data, to crosscheck conflicting

<https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>

cases, to validate trustworthy relationships, to disclose inherent clusters, and to uncover hidden relationships and models.

5. Discussion This paper provides a current outlook on the subject of Big Data. From the ideas presented here it is clear that Big Data has the potential to become the next biggest chapter if you will in the IT industry. However it will take time before we see the widespread use of this technology and truly uncover its full potentials. And it is fair to say that Big Data will be a technology that will be widely used by everyone in the every aspects of society. 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. Today's analysis are impeded by a tedious process of exporting data from the database, performing a non-SQL process and bringing the data back. This is an obstacle to carrying over the interactive elegance of the first generation of SQL-driven OLAP systems into the data mining type of analysis that is in increasing demand. A tight coupling between declarative query languages and the functions of such packages will benefit both expressiveness and performance of the analysis.

6. Using Big Data When a company is looking to step into the Big Data technology, they will be faced with questions such as given below. - - What is the cost? - How to integrate with business? - We have entered an era of Big Data. Through better analysis of the large volumes of data that are becoming available [8], there is the potential for making faster advances in many scientific disciplines and improving the profitability and success of many enterprises. However, many technical challenges described in this paper must be addressed before this potential can be realized fully. The

challenges include not just the obvious 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, and will not be addressed naturally by the next generation of industrial products. We must support and encourage fundamental research towards addressing these technical challenges if we are to achieve the promised benefits of Big Data. What are the vendor platforms? - 4. Future of Data-Driven Innovation How to build the infrastructure? How to capture business value? I believe that the discussion in this paper would have been able to answer the some of these questions. Also we understand the demand for Big Data technology is becoming increasingly large, however companies need to acquire more knowledge in both technical tasks and data management tasks. This will create a large demand for data science experts and IT experts of Big Data technologies.

Acknowledgment I wish to thank my supervisor Mr. Saminda Premaratne who provided guidance and valuable advice. I also show my gratitude to my friends and all who contributed in one way or the other in completing this paper.

References [1] Computing Research Association, Big Data White Paper. But even greater than the challenges are the opportunities that big data presents. McKinsey calls big 3 [2] Challenges and Opportunities with Big Data, A community white paper developed by leading researchers across the United States [5] IBM Hurwitz & Associates, Fern Halper, January 2012, Four Vendor Views on Big Data and Big Data Analytics. [3] James Manyika, <https://assignbuster.com/databig-data-and-future-of-data-driven-innovation/>



Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. " Big data: The next frontier for innovation, competition, and productivity." McKinsey Global Institute. May 2011. [6] Pattern-Based Strategy: Getting Value from Big Data. Gartner Group press release. July 2011. [7] The Data Warehousing Institute, Big Data Analytics White Papers. [8] The Age of Big Data. Times, Feb 11, 2012. [4] McKinsey Global Institute, the next frontier for innovation, competition and productivity [http://goo. gl/Ug9ze](http://goo.gl/Ug9ze) 4 Steve Lohr. New York