

## Chapter 2

# Statistical and Computational Needs for Big Data Challenges

**Soraya Sedkaoui**

*Khemis Miliana University, Algeria & Montpellier University, France*

### **ABSTRACT**

*The traditional way of formatting information from transactional systems to make them available for “statistical processing” does not work in a situation where data is arriving in huge volumes from diverse sources, and where even the formats could be changing. Faced with this volume and diversification, it is essential to develop techniques to make best use of all of these stocks in order to extract the maximum amount of information and knowledge. Traditional analysis methods have been based largely on the assumption that statisticians can work with data within the confines of their own computing environment. But the growth of the amounts of data is changing that paradigm, especially which ride of the progress in computational data analysis. This chapter builds upon sources but also goes further in the examination to answer this question: What needs to be done in this area to deal with big data challenges?*

DOI: 10.4018/978-1-5225-3203-3.ch002

## **INTRODUCTION**

With the advent of digital technology and smart devices, a large amount of digital data is being generated every day. Individuals are putting more and more publicly available data on the web. Many companies collect information on their clients and their respective behavior. As such, many industrial and commercial processes are being controlled by computers. The results of medical tests are also being retained for analysis. Financial institutions, companies, and health service providers, administrations generate large quantities of data through their interactions with suppliers, patients, customers, and employees. Beyond those interactions, large volumes of data are created through Internet searches, social networks, GPS systems, and stock market transactions.

This brings us to think about the legend of the wise ‘Sissa’ in India. When King ‘Belkib’ asked about the reward he desired, after his invention, he asked to receive a grain of rice for the first square, two grains for the second, four grains for the third and so on. The king agreed, but he didn’t know that on the last square of the board he should drop  $2^{63}$  grains, or more than 700 billion tons. In their book “Race Against the Machine,” Brynjolfsson and McAfee (2011) referenced the fable of the chess and rice grains to make the point that “exponential increases initially look a lot like linear, but they are not. As time goes by – as the world move into the second half of the chessboard – exponential growth confounds our intuition and expectation”.

Thus currently, not only is the quantity of digitally stored data much larger, but the type of data is also very varied, thanks to the various new technologies (Sedkaoui & Monino, 2016). Data volume will continue to grow and in a very real way, the data produced, as well as other data accumulated, constitutes a constant source of knowledge. This widespread production of data has resulted in the ‘data revolution’ or the age of ‘big data’. Big data gets global attention and can be best described using the three Vs: volume, variety and velocity. These three dimensions often are employed to describe the phenomenon. Each dimension presents both challenges for data management and opportunities to advance decision-making. In another way, every data tells a story and data analytics, in particular the statistical methods coupled with the development of IT tools, piece together that story’s reveal the underlying message.

This 3 V’s provide a challenge associated with working with big data. The volume put the accent on the storage, memory and computes capacity of a computing system and requires access to a computing cloud. Velocity stresses the rate at which data can be absorbed and meaningful answers produced. The variety makes it difficult to develop algorithms and tools that can address that large variety of input data. So, there are still many difficulties and challenges in the use of big data technologies. And, if decision-makers can’t understand the power of data processing and analytics, they

31 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/statistical-and-computational-needs-for-big-data-challenges/202913](http://www.igi-global.com/chapter/statistical-and-computational-needs-for-big-data-challenges/202913)

## Related Content

---

### Phase Unwrapping Using Energy Minimization Methods for MRI Phase Image

Kusworo Adi, Tati L. R. Mengko, Andriyan B. Suksmono and H. Gunawan (2012). *Emerging Communication Technologies for E-Health and Medicine* (pp. 252-258). [www.irma-international.org/chapter/phase-unwrapping-using-energy-minimization/65718](http://www.irma-international.org/chapter/phase-unwrapping-using-energy-minimization/65718)

### Inhibitors of Physicians' Use of Mandatory Hospital Information Systems (HIS)

Heiko Gewald and Corinna Gewald (2018). *International Journal of Healthcare Information Systems and Informatics* (pp. 29-44). [www.irma-international.org/article/inhibitors-of-physicians-use-of-mandatory-hospital-information-systems-his/192185](http://www.irma-international.org/article/inhibitors-of-physicians-use-of-mandatory-hospital-information-systems-his/192185)

### Response to Intervention: Assistive Technologies which can Help Teachers with Intervention Programming and Assessment

Michael W. Dunn (2010). *Handbook of Research on Human Cognition and Assistive Technology: Design, Accessibility and Transdisciplinary Perspectives* (pp. 313-324). [www.irma-international.org/chapter/response-intervention-assistive-technologies-can/42846](http://www.irma-international.org/chapter/response-intervention-assistive-technologies-can/42846)

### Monitoring and Assisting Maternity-Infant Care in Rural Areas (MAMICare)

Juan C. Lavariega, Gustavo A. Córdova, Lorena G. Gómez and Alfonso Avila (2014). *International Journal of Healthcare Information Systems and Informatics* (pp. 32-43). [www.irma-international.org/article/monitoring-and-assisting-maternity-infant-care-in-rural-areas-mamicare/124118](http://www.irma-international.org/article/monitoring-and-assisting-maternity-infant-care-in-rural-areas-mamicare/124118)

### Implementation of a Computerized System in an Oncology Unit

Taxiarchis Botsis and Konstantinos Syrigos (2007). *International Journal of Healthcare Information Systems and Informatics* (pp. 31-38). [www.irma-international.org/article/implementation-computerized-system-oncology-unit/2209](http://www.irma-international.org/article/implementation-computerized-system-oncology-unit/2209)