

Chapter 68

The Impact of Big Data on Security

Mohammad Alaa Hussain Al-Hamami
Applied Science University, Bahrain

ABSTRACT

Big Data is comprised systems, to remain competitive by techniques emerging due to Big Data. Big Data includes structured data, semi-structured and unstructured. Structured data are those data formatted for use in a database management system. Semi-structured and unstructured data include all types of unformatted data including multimedia and social media content. Among practitioners and applied researchers, the reaction to data available through blogs, Twitter, Facebook, or other social media can be described as a “data rush” promising new insights about consumers’ choices and behavior and many other issues. In the past Big Data has been used just by very large organizations, governments and large enterprises that have the ability to create its own infrastructure for hosting and mining large amounts of data. This chapter will show the requirements for the Big Data environments to be protected using the same rigorous security strategies applied to traditional database systems.

INTRODUCTION

The term Big Data has a relative meaning and tends to denote bigger and bigger data sets over time. In computer science, it refers to data sets that are too big to be handled by regular storage and processing infrastructures. It is evident that large datasets have to be handled differently than small ones; they require different means of discovering patterns, or sometimes allow analyses that would be impossible on a small scale.

In the social sciences and humanities as well as applied fields in business, the size of data sets thus tends to challenge researchers as well as software or hardware. This may be especially an issue for disciplines or applied fields that are more or less unfamiliar with quantitative analysis. (Manovich, 2012) sees knowledge of computer science and quantitative data analysis as a determinant for what a group of researchers will be able to study. He fears a “data analysis divide” between those equipped with the necessary analytical training and tools to actually make use of the new data sets and those who will inevitably only be able to scratch the surface of this data.

DOI: 10.4018/978-1-4666-9840-6.ch068

New analytical tools tend to shape and direct scholars' ways of thinking and approaching their data. The focus on data analysis in the study of Big Data has even led some to the assumption that advanced analytical techniques make theories obsolete in the research process. Research interest could thus be almost completely steered by the data itself. But being driven by what is possible with the data may cause a researcher to disregard vital aspects of a given research object. (Boyd and Crawford, 2012) underline the importance of the (social) context of data that has to be taken into account in its analysis. The scholars illustrate how analyses of large numbers of messages ("tweets") from the micro blogging service.

Twitter currently used to describe aggregated moods or trending topics—without researchers really discussing what and particularly who these tweets represent: Only parts of a given population are even using Twitter, often in very different ways. As Boyd and Crawford point out, these contexts are typically unknown to researchers who work with samples of messages captured through Twitter. In addition, Big Data analyses tend only to show what users do, but not why they do it. In his discussion of tools for Big Data analysis, (Manovich, 2012) questions the significance of the subsequent results in terms of their relevance for the individual or society. The issue of meaning of the observed data and/or analyses is thus of vital importance to the debate around Big Data.

Many organizations are relying on multiple tools to produce the necessary security data. This will lead to a huge and a complex data feeds that must be analyzed, normalized, and prioritized. The scale of security data that needs analysis has simply become too big and complex to handle, and its takes a very long time. According to the Verizon 2013 Data Breach Investigations Report, 69% of breaches were discovered by a third party and not through internal resources and this is an example to clarify why the role of Big Data is important in security.

"BIG DATA" TERM

"Big Data" is a term that has quickly achieved widespread use among technologists, researchers, the media and politicians. Perhaps due to the speed of dissemination the use of the term has been rather nebulous in nature. The concept of Big Data can be framed by one of three perspectives. The first is a response to the technology problems associated with storing, securing and analyzing the ever-increasing volumes of data being gathered by organizations. This includes a range of technical innovations, such as new types of database and 'cloud' storage that enable forms of analysis that would not previously have been cost effective. The second perspective focuses on the commercial value that can be added to organizations through generating more effective insights from this data. This has emerged through a combination of better technology and greater willingness by consumers to share personal information through web services. The third perspective considers the wider societal impacts of Big Data, particularly the implications for individual privacy, and the effect on regulation and guidelines for ethical commercial use of this data. We now consider each of these perspectives on Big Data in more detail.

BIG DATA AND TECHNOLOGY INNOVATION

In its original form, Big Data referred to technical issues relating to the large volumes of data being created. While the rate at which data has been generated by information technology has always been increasing, recent growth produces some startling statistics (Qin and Li, 2013). Take the following two examples:

22 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/the-impact-of-big-data-on-security/150227

Related Content

An Efficient Pruning and Filtering Strategy to Mine Partial Periodic Patterns from a Sequence of Event Sets

Kung-Jiuan Yang, Tzung-Pei Hong, Yuh-Min Chen and Guo-Cheng Lan (2014). *International Journal of Data Warehousing and Mining* (pp. 18-38).

www.irma-international.org/article/an-efficient-pruning-and-filtering-strategy-to-mine-partial-periodic-patterns-from-a-sequence-of-event-sets/110384

Association Rule Mining in Developmental Psychology

D. A. Nembhard, K. K. Yip and C. A. Stifter (2013). *Data Mining: Concepts, Methodologies, Tools, and Applications* (pp. 1737-1751).

www.irma-international.org/chapter/association-rule-mining-developmental-psychology/73520

An Intelligent Heart Disease Prediction Framework Using Machine Learning and Deep Learning Techniques

Nasser Allheeb, Summrina Kanwal and Sultan Alamri (2023). *International Journal of Data Warehousing and Mining* (pp. 1-24).

www.irma-international.org/article/an-intelligent-heart-disease-prediction-framework-using-machine-learning-and-deep-learning-techniques/333862

A Method for Generating Comparison Tables From the Semantic Web

Arnaud Giacometti, Béatrice Markhoff and Arnaud Soulet (2022). *International Journal of Data Warehousing and Mining* (pp. 1-20).

www.irma-international.org/article/a-method-for-generating-comparison-tables-from-the-semantic-web/298008

Emotion-Drive Interpretable Fake News Detection

Xiaoyi Ge, Mingshu Zhang, Xu An Wang, Jia Liu and Bin Wei (2022). *International Journal of Data Warehousing and Mining* (pp. 1-17).

www.irma-international.org/article/emotion-drive-interpretable-fake-news-detection/314585