Big Data Issues:
Analytics and Security

B

Dina Darwish
Ahram Canadian University, Egypt

INTRODUCTION

Big data is a collection of structured, semi-structured, and unstructured data that is gathered by businesses and may be mined to be used in advanced analytics applications using machine learning and predictive modelling. Big data is used by businesses to enhance operations, deliver better customer service, develop individualized marketing campaigns, and carry out other tasks that can ultimately boost sales and profits. Because they can act more quickly and with greater knowledge, businesses who use it efficiently may have a competitive advantage over those that don't. Data scientists and other data analysts need to have a thorough comprehension of the available data and a clear understanding of what they're looking for in it in order to provide reliable and pertinent results from big data analytics applications. As a result, a vital preparatory stage in the analytics process is data preparation, which involves profiling, cleansing, validation, and transformation of data sets. Securing Big Data, which frequently contains sensitive information, is another crucial concern that needs to be addressed.

This chapter is going to address two important issues for Big Data; which are analytics and security. Since, today companies and data scientists work on Big Data analytics to retrieve relevant information, and provide essential statistics to help companies and research labs in finding solutions to some questions, and improve their sales and competitiveness in the market, and at the same time, keeping these information secured. The objectives of this article are:

1. Concentrating on Big Data security related issues
2. Presenting various solutions to Big Data security challenges
3. Clarifying importance of Big Data analytics, especially for companies
4. Focusing on several solutions to issues related to Big Data security
5. Providing strategies in general for managing effectively Big Data
6. Discussing conclusion, and future research directions.

This chapter is organized as follows; the background represents the second section, where literature review is mentioned, the third section is composed of focus of the chapter, which is definition of Big Data, its importance, applications, and analytics, section 4 focuses on solutions and recommendations for Big Data security and analytics challenges, and finally, comes conclusion and future research directions.

BACKGROUND

Research needs to be done on two crucial Big Data issues: analytics and security. Some research on these two issues are mentioned in the following sections. Social networks, such as Facebook and Twit-

DOI: 10.4018/978-1-6684-7366-5.ch020
This article, published as an Open Access article in the gold Open Access encyclopedia, Encyclopedia of Information Science and Technology, Sixth Edition, is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

ter are major producers of Big Data. Social networks like Facebook and Twitter are leading producers of big data. Social Network topology has strong impact on physical technological networks as most of the traffic is contributed by these social network sites and related ones. Some approaches and examples of the use of social network analysis in the design of technology networks and vice versa are explored in research (Cheng et al., 2013). A user can be a member of multiple networks at the same time, and these networks can combine to produce a composite social network in which the person's activity varies across networks. Moreover, the user may share similar latent interests with other users throughout these networks. E. Zhong (Zhong et al.,2012) proposed a model for adaptive transfer of knowledge from composite social networks to forecast human behavior for use in social marketing, service suggestions, and personalization. Using personal ad hoc clouds of users in social networks to address big data processing difficulties by leveraging the social network paradigm for generating information from big data was discussed in research (Tan et al., 2013). The combination of IoT, big data analytics, and complicated event processing techniques is suggested by authors (Tasweef et al., 2015) as a solution to the main problems with handling data in the healthcare industry. They proposed an all-encompassing healthcare system that could carry out tasks like drug detection, monitor patients from a distance, help with health insurance settlement, and advance the therapeutic outcomes.

Through the use of big data analytics, organizations can obtain useful data and patterns that could have an impact on their operations (Gandomi & Haider, 2015). Therefore, in order to determine the relationships between features and predict future observations, extensive data analysis is required. Big data analytics refers to methods used to draw conclusions from enormous databases (Labrinidis & Jagadish, 2012). The outcomes of big data analytics can enhance decision-making and boost organizational effectiveness. In order to extract knowledge from the data, many analytical techniques are developed, such as; Descriptive analytics is focused with examining historical data of a business to describe what happened in the past; (Joseph & Johnson, 2013), To forecast potential outcomes, predictive analytics focuses on various statistical modelling and machine learning techniques (Waller & Fawcett, 2013). Descriptive and predictive analytics are combined in prescriptive analytics to suggest the best action for enhancing company procedures (Joseph & Johnson, 2013).

Vaishya and his colleagues (Vaishya et al., 2020) explored the primary uses of AI for preventing and combating Coronavirus Disease through the use of big data analytics (COVID-19). The scientists identified seven uses of AI for the COVID-19 pandemic, including: disease detection, patient treatment monitoring, contact tracing, cases and deaths prediction, medicine manufacture, workload reduction, and disease prevention. The following, however, are factors that this paper neglects to consider: the small number of papers examined; the lack of a statement regarding the study selection procedure; and the absence of any qualitative characteristics. Furthermore, a thorough taxonomy based on AI methods was not provided. Finally, Pham and his colleagues (Pham et al., 2020) described how to organize and evaluate the enormous volume of data obtained by the COVID-19 disease using big data and AI techniques. Five categories—including COVID-19 outbreak prediction, viral tracking, diagnosis and treatment, and drug discovery—are taken into consideration when evaluating certain big data tools. The associated problems with the solutions under examination were then highlighted.

Using machine learning techniques, Kaur and his colleagues (Kaur et al., 2018) suggested a unique approach for intelligent healthcare information systems. There are four layers in the suggested model. Diverse data sources are handled by the data source layer. The storage optimization procedure is controlled by the data storage layer. To make the best use of system resources, a number of techniques have been applied, including normalization and indexing. The data security layer uses a variety of data security and privacy techniques, including data masking, granular control over data access, activity monitor-

21 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/big-data-issues/319015

Related Content

Coping Strategies and Paradoxes Related to BYOD Information Security Threats in France
Paméla Bailletteand Yves Barlette (2020). *Journal of Global Information Management (pp. 1-28).*
www.irma-international.org/article/coping-strategies-and-paradoxes-related-to-byod-information-security-threats-in-france/246094

Transnational Information Systems: Development and Management Issues
Paul Mantelaersand Wander van den Berg (2002). *Advanced Topics in Global Information Management, Volume 1 (pp. 120-135).*
www.irma-international.org/chapter/transnational-information-systems/4493

Misaligned Market: The Importance of Industry Context in Technology-Mediated exchanges
Ruey-Lin Hsiao (2007). *Journal of Global Information Management (pp. 69-87).*
www.irma-international.org/article/misaligned-market-importance-industry-context/3657

Research, Application, and Innovation of LiDAR Technology in Spatial Archeology
Carolina Collaroand Martin Herkommer (2025). *Encyclopedia of Information Science and Technology, Sixth Edition (pp. 1-33).*
www.irma-international.org/chapter/research-application-and-innovation-of-lidar-technology-in-spatial-archeology/324093

Does Strategic Outsourcing Undermine the Innovative Capability of Organizations?
A. Hoecstand P. Trott (2008). *Global Information Technologies: Concepts, Methodologies, Tools, and Applications (pp. 3478-3502).*
www.irma-international.org/chapter/does-strategic-outsourcing-undermine-innovative/19193