Chapter 6
Big Data Analytics:
An Expedition Through Rapidly Budding Data Exhaustive Era

Sreenu G.
Muthoot Institute of Technology and Science, India

M.A. Saleem Durai
VIT University, India

ABSTRACT

Advances in recent hardware technology have permitted to document transactions and other pieces of information of everyday life at an express pace. In addition of speed up and storage capacity, real-life perceptions tend to transform over time. However, there are so much prospective and highly functional values unseen in the vast volume of data. For this kind of applications conventional data mining is not suitable, so they should be tuned and changed or designed with new algorithms. Big data computing is inflowing to the category of most hopeful technologies that shows the way to new ways of thinking and decision making. This epoch of big data helps users to take benefit out of all available data to gain more precise systematic results or determine latent information, and then make best possible decisions. Depiction from a broad set of workloads, the author establishes a set of classifying measures based on the storage architecture, processing types, processing techniques and the tools and technologies used.

INTRODUCTION

Big data is a latest happening with the potential to trans-form the values of products and services in industry and business. A definition for big data is given by ("Big Data," n. d.) as "Big data is high-volume, high-velocity, and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

Rapid advancements in technical knowledge and networking lead to tremendous growth of information in all fields such as education, health science and business. The materialization of novel technologies such as Internet, Network-of-Things and large scale wireless sensor systems facilitates the gathering of data from an increasing volume, velocity and variety of networked sensors for analysis. Data volume corresponds to the magnitude of data that can be warehoused and evaluated. Data velocity symbolizes the speed of data aggregation and streaming. Data variety point towards various sources of data's such as images, audio, text, video, etc. Veracity refers to disorderliness or reliability of data.

The progresses of computer systems as well as internet technologies as per Moore's law witnessed the fact that the difficulties of managing the extensive data still occur at the age of big data. Heterogeneous nature of data leads to the classification such as structured and unstructured data. The fact that science is lagging behind the real world in the proficiencies of handling large volume of data and realizing useful facts from immense volume of data. Techniques to develop architecture capable of handling massive volume of data comprise various hardware capabilities and various programming models.

The enormous growth of data needs enhancement from old-fashioned data processing solutions to systems which handles a constant stream of real time data. Several distributed stream processing engines are also taken for a review. Applications which need valuable analyses of huge datasets are widely increasing nowadays. As a result, data evolution will primarily outperform expected improvements in the cost and concentration of storage technologies, the reachable computational power for processing it, and the connected energy path. Discussion of these methods is described in following section of this chapter. The data processing section will analyze various mining techniques developed for processing big data. Various tools used to make the big data processing easier will be discussed.

ORGANIZATION OF BIG DATA

The way of capturing and storing data has been changed by the arrival of big data. Disk access latencies are reduced by the introduction of storage technologies like

Related Content

Assessing the Impact of Governmental Regulations on the IT Industry: A Neo Institutional Theory Perspective
Sushma Mishraand Amita Goyal Chin (2008). *Computer Security, Privacy and Politics: Current Issues, Challenges and Solutions (pp. 36-54).*
www.irma-international.org/chapter/assessing-impact-governmental-regulations-industry/6860

Digital Audio Watermarking
Changsheng Xuand Qi Tian (2005). *Multimedia Security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property (pp. 126-156).*
www.irma-international.org/chapter/digital-audio-watermarking/27047

DDoS Attack Simulation and Machine Learning-Based Detection Approach in Internet of Things Experimental Environment
Hongsong Chen, Caixia Mengand Jingjiu Chen (2021). *International Journal of Information Security and Privacy (pp. 1-18).*
www.irma-international.org/article/ddos-attack-simulation-and-machine-learning-based-detection-approach-in-internet-of-things-experimental-environment/281038

Application of Representation Learning-Based Chronological Modeling for Network Intrusion Detection
Nitin O. Mathur, Chengcheng Li, Bilal Gonenand Kijung Lee (2022). *International Journal of Information Security and Privacy (pp. 1-32).*
www.irma-international.org/article/application-of-representation-learning-based-chronological-modeling-for-network-intrusion-detection/291701

Privacy Preserving in the Modern Era: A Review of the State of the Art
Rihab Boussada, Mohamed Elhoucine Elhdhili, Balkis Hamdaneand Leila Azouz Saidane (2022). *Handbook of Research on Technical, Privacy, and Security Challenges in a Modern World (pp. 1-24).*
www.irma-international.org/chapter/privacy-preserving-in-the-modern-era/312413