Emerging Tools and Technologies in Data Science

Mahmud Ullah
https://orcid.org/0000-0001-7472-2477
Department of Marketing, University of Dhaka, Bangladesh

Nayem Rahman
Anderson College of Business and Computing, Regis University, USA

INTRODUCTION

Necessary and reasonable use of data to produce authentic and effective information to run different government, nongovernment, nonprofit organizations, and mainly large size businesses had always been there up to certain extent. Every phase of industrial revolution had added new dimensions to increase the size and frequencies of data usage, making it unmanageable at times to deal with the required data by the existing knowledge of the time.

Consequently, every phase of industrial revolution gave birth of many new fields of knowledge related to dealing with data or data management, which may be collectively termed as 'Data Science'. After the invention of information technology, now in the rapidly growing current phase of industrial revolution, which is technically termed by the professionals, and popularly known to people as Fourth Industrial Revolution, referred as Industry 4.0 in brief as well, every single entity in the world is immersed in the large oceans of data, data, and data only, irrespective of any entity's consent to be so.

Hence it is becoming even more & more difficult every day to manage the huge volume of data we have to deal with as an individual, or as a small or large scale business, or as government, nongovernment, and nonprofit organizations. The term 'Big Data' has evolved to refer to the sheer size of data, and the scope, importance, usefulness etc. of these data we have deal with in our everyday lives nowadays.

Luckily, new knowledge on tools and technologies to handle these big data, an obvious reality of this industry 4.0 era, is also evolving simultaneously. The big question is, how useful this continuous addition & learning of new knowledge in the field of data science are, in terms of wide spread availability, easy accessibility, smooth understanding, price structure, cost saving / cutting, effectiveness, efficiency, user friendliness etc., and how to make the optimum use of the ever increasingly emerging knowledge on tools and technologies in the field of data science, the most important field for every single country, organization, and individual around the world.

This is the most desirable field of knowledge as well, mainly to the younger generation everywhere around the world. Everybody would love to have some expertise at least in dealing with data scientifically using the most up to date tools and technologies of data management, which is virtually pushing the field of data science to flourish more rapidly than ever before.

In this context, the authors prepared this paper covering the nature of ever increasing evolution of overwhelming knowledge in the field of data science, focusing mainly on the discovery or emergence of tools and technologies to deal with data, on the basis of their experience and expertise of doing or being involved in similar kind of multiple research projects or activities in small scale in the field of data

DOI: 10.4018/978-1-7998-9220-5.ch099

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

science. They believe that this paper would be worth reading for the professionals and academicians to have an in-depth idea of the continuing development in the field of data science, and choose the best possible tools or technologies for their respective organizations corresponding to their specific purposes.

BACKGROUND

The advent of computing technologies, software engineering, data warehousing technologies, cloud computing, etc., and very rapid emergence and adoption of smart-phone around the globe, have made modern day business environment much more competitive than ever before. To survive and sustain facing the immense competition due to all these fast emerging technologies, businesses have to make informed decision by analyzing business data using the most effective, efficient, and up to date data dealing tool, which will ultimately ensure the use of the least expensive data handling procedure. This paper will try to identify the most economic ways of using data handling tools and technologies.

Data analyses are mainly done using data stored in the enterprise data warehouses, which (data) are mainly created in, and collected from operational data stores. Everybody wants to have faster access to information but not quite eager to pay as much. Business enterprises struggle continuously to make a balance between these two conflicting expectations. This paper aims to identify the best possible ways to make this balance, e.g. switching to cloud computing from maintaining enterprises' own data warehouse.

Handling huge data volume, which is referred as big data, and coming from different sources like social media, videos, mobile phone etc., is a real challenge. Fortunately, new tools and parallel processing database engines are now available to assist with overcoming that challenge. This paper attempts to identify the existing and emerging technologies of data storage, retrieval, parallel processing etc. with as much details as possible to help people choose the most suitable one for any person, or organization to handle these large data.

Thus, this chapter aims to make an exhaustive list of all the existing and emerging technologies to deal with data of any volume - big, large, or huge, by displaying the required details in a systematic manner classifying and categorizing those, making those easy to comprehend and compare. Exhaustive literature review has been done to start with. Information has been collected from all the stakeholders involved in producing, marketing, and using data handling / managing / interpreting tools and technologies to generate authentic information by systematically storing, instantly retrieving, and objectively analyzing different types of data of different volume, mostly large or huge referred as big data.

Global nature of modern business, triggered and facilitated mainly by the drastic development in the ever changing digital devices and software engineering has made the business environment much more competitive, and consequently kind of volatile in many senses, especially in this era of industry 4.0, evolving with an extraordinary pace unlike the other phases of industrial revolution. Making informed decisions on the basis of authentic data and information has hence become obvious for the organizations to survive, sustain, and succeed.

Companies typically create and collect data in operational data stores. This data is then moved to the enterprise data warehouse for various analytical needs. There is a push to access information faster, often as soon as it is created. To achieve that, database engines need to have a faster processing capability (Rahman, 2007). While organizations want a state of the art data warehousing and Business Intelligence (BI) environment, they also continuously strive to bring down their IT infrastructure budget. Data warehouse and BI tools involve huge capital expenditure. Small and medium-sized companies cannot afford that. With the advent of cloud computing data warehouse infrastructure costs could be avoided as

13 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/emerging-tools-and-technologies-in-data-science/317576

Related Content

Recommendation System: A New Approach to Recommend Potential Profile Using AHP Method
Safia Baali (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-14).*
www.irma-international.org/article/recommendation-system/279278

Wind Speed Prediction at Multiple Heights for Offshore Wind Turbines Using Deep Learning
Dipankar Dutta (2025). *Advanced Interdisciplinary Applications of Deep Learning for Data Science (pp. 111-142).*
www.irma-international.org/chapter/wind-speed-prediction-at-multiple-heights-for-offshore-wind-turbines-using-deep-learning/366146

Decision-Making Systems
Ghalia Nasserddineand Amal A. El Arid (2023). *Encyclopedia of Data Science and Machine Learning (pp. 1391-1407).*
www.irma-international.org/chapter/decision-making-systems/317548

An Overview of Available Deepfake Datasets in Neural Network-Based Soil and Weather Prediction Models for High Quality Crops
A. Anisha Sanjeetha, R. Sivarajand P. Uma (2023). *Handbook of Research on Advanced Practical Approaches to Deepfake Detection and Applications (pp. 55-71).*
www.irma-international.org/chapter/an-overview-of-available-deepfake-datasets-in-neural-network-based-soil-and-weather-prediction-models-for-high-quality-crops/316743

Intelligent System for Credit Risk Management in Financial Institutions
Philip Sarfo-Manu, Gifty Siawand Peter Appiahene (2019). *International Journal of Artificial Intelligence and Machine Learning (pp. 57-67).*
www.irma-international.org/article/intelligent-system-for-credit-risk-management-in-financial-institutions/238128