Chapter 29
Big Data Analytics Tools and Platform in Big Data Landscape

Mohd Imran
Aligarh Muslim University, India

Mohd Vasim Ahamad
Aligarh Muslim University, India

Misbahul Haque
Aligarh Muslim University, India

Mohd Shoaib
Aligarh Muslim University, India

ABSTRACT

The term big data analytics refers to mining and analyzing of the voluminous amount of data in big data by using various tools and platforms. Some of the popular tools are Apache Hadoop, Apache Spark, HBase, Storm, Grid Gain, HPCC, Casandra, Pig, Hive, and No SQL, etc. These tools are used depending on the parameter taken for big data analysis. So, we need a comparative analysis of such analytical tools to choose best and simpler way of analysis to gain more optimal throughput and efficient mining. This chapter contributes to a comparative study of big data analytics tools based on different aspects such as their functionality, pros, and cons based on characteristics that can be used to determine the best and most efficient among them. Through the comparative study, people are capable of using such tools in a more efficient way.

INTRODUCTION

Big data technology is a revolutionary technology which is currently adapted by all scale organization varying from small private industries to large government organization. It is now agreeable among all academicians and entrepreneurs that big data is having some game changer capabilities which makes the big data analytics a great and powerful tool for market research. Now at these times, most of the

DOI: 10.4018/978-1-6684-3662-2.ch029

Copyright © 2022, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

business hubs as well as small organizations are coming forward to opt the big data analytics tool to dig up their strategy of marketing and produce maximum output from same; despite of having their forefront challenges of investment and cost of marinating stabilization. Another benefit which business organizations are looking into consists of a well new customer experience, more reliable and efficient goal and a better look up of organization from a completely different perspective. The extensive use of this technology is in educational purpose as well as in health care organization. Research, which is augmented educational wing uses the tools of big data analytics at various level, and for numerous applications (Samiya, kashish & Alam, 2016). According to study, educational as well as healthcare sector is generating huge amount of data which make it a potential source for big data analytics but to make it happen all the data must be refined, recorded and managed. Big data analytics also have its security concern and challenges (Jayasingh et. al). The analytics tools used in big data for security must have their encrypting capabilities to protect large exploding amount of data at every level from system to forensic level. To enhance the performance, reliability, and accuracy of system, people should know the environment where the analytics tool is most suited. The Apache Hadoop is one the revolutionary platform that provides various remarkable analytics tool to manage processing and handling. Some of these are suitable for collaborative distributed computing, some are well adopted for real time streaming and likewise some are popular for their graph representing capability. Apache Hadoop consists of various tools which can be categorized as Business tools, data Science tool, Interaction tool, Sql/NoSql tool, Cognition, conversion, security, search and storage tool. Apache Spark is another add on for big data application which provides processing speed faster than Hadoop is nearly hundred times faster. It is a remarkable analytical tool, well known for its distributed computing and graph computational analytics. There are some other analytical tools are available like Hive, Pig, HBase, Cassandra, Storm, HPCC which are handy. In order to exploit feature of these tools, we have to learn a comparative analysis, by exposing them in different factors and parameters. By learning this, People are capable of using them in more simplified manner.

BACKGROUND AND MAIN FOCUS

The Big data analytics is new trending analytical standard used to fetch previously collected data which is generated by numerous applications for pattern searching that cannot be examined, processed, managed and categorized by any other existing tools or technologies (Yadav, Verma, and Kaushik, 2015).Hence new technology or tools must be adapted which can handle vast datasets generated from commodity servers which are distributed all across the globe. It is a technique of extracting useful correlated informations form massive dataset. Big data can be categorized in structured, Semi and Unstructured format. Mining of these structured, unstructured, and unrelated information collected from vast corporations, research, and healthcare organizations and make it useful by managing, structuring, controlling is main objective of big data analytics. Together, Big data analytics (BDA) is information managing tool that uncover the hidden pattern, correlated from vast big data set to make a decision control for large organization for optimized performance. The main focus is to deploy big data analytics tool in different sector of market, in order to obtain various pattern of market research.

- **Big Data in Healthcare Sector:** The health sectors are applying the big data analytics for determining the pattern of disease in various patients as well as in demographic variations. The digital

8 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-analytics-tools-and-platform-in-big-data-landscape/291004

Related Content

EMG-Based Mobile Assessment System for Neck and Shoulder Fatigue
Pei Lun Lai, Hsiu-Sen Chiang and Qi-An Huang (2017). *International Journal of Big Data and Analytics in Healthcare (pp. 39-50).*
www.irma-international.org/article/emg-based-mobile-assessment-system-for-neck-and-shoulder-fatigue/204447

Introduction to Programming R and Python Languages
(2017). *Comparative Approaches to Using R and Python for Statistical Data Analysis (pp. 32-77).*
www.irma-international.org/chapter/introduction-to-programming-r-and-python-languages/175144

Proximate Breast Cancer Factors Using Data Mining Classification Techniques
Alice Constance Mensah and Isaac Ofori Asare (2019). *International Journal of Big Data and Analytics in Healthcare (pp. 47-56).*
www.irma-international.org/article/proximate-breast-cancer-factors-using-data-mining-classification-techniques/232335

Convolutional Locality-Sensitive Dictionary Learning for Facial Expressions Detection
Benjamin Ghansah (2022). *International Journal of Data Analytics (pp. 1-28).*
www.irma-international.org/article/convolutional-locality-sensitive-dictionary-learning-for-facial-expressions-detection/297520

Towards Clinical and Operational Efficiency through Healthcare Process Analytics
Vassiliki Koufi, Flora Malamateniou and George Vassilacopoulos (2016). *International Journal of Big Data and Analytics in Healthcare (pp. 1-17).*
www.irma-international.org/article/towards-clinical-and-operational-efficiency-through-healthcare-process-analytics/171401