Chapter 11
Amelioration of Big Data Analytics by Employing Big Data Tools and Techniques

Stephen Dass
Vellore Institute of Technology, India

Prabhu J.
Vellore Institute of Technology, India

ABSTRACT

This chapter describes how in the digital data era, a large volume of data became accessible to data science engineers. With the reckless growth in networking, communication, storage, and data collection capability, the Big Data science is quickly growing in each engineering and science domain. This paper aims to study many numbers of the various analytics ways and tools which might be practiced to Big Data. The important deportment in this paper is step by step process to handle the large volume and variety of data expeditiously. The rapidly evolving big data tools and Platforms have given rise to numerous technologies to influence completely different Big Data portfolio.In this paper, we debate in an elaborate manner about analyzing tools, processing tools and querying tools for Big datahese tools used for data analysis Big Data tools utilize numerous tasks, like Data capture, storage, classification, sharing, analysis, transfer, search, image, and deciding which might also apply to Big data.

INTRODUCTION

Current advancement in the field digital information improves the data which are exceptional to both software and hardware. About 70% unstructured data deals with multimedia data, in that 60% of them are from internet traffic (Boyd & Crawford, 2012; Hartmann et al., 2014; Jagadish et al., 2014; Katal, Wazid, & Goudar, 2013; Purcell, 2013). Unexpectedly huge data creates stints multi-media data semantic definitions searched by conventional methods are difficult to any set of forms. Unsorted raw data are complicated to deal directly so few easy and machine processing forms are made to design semantic data. This type of data works on content-based retrieval methods from which data are restored. This phenomenon is known as Feature Extraction (Katal, Wazid, & Goudar, 2013). Miloslavaskaya and Tolstoy

(2014) state "…big data concept are the datasets of such size and structure that exceed the capabilities of traditional programming tools (databases, software, etc.) for data collection, storage and processing in a reasonable time and a-fortiori exceed the capacity of their perception by a human…"

In General, Big Data is exported as data wealth peculiarize as high volume, velocity, and variety to get particular technology and analytical methods to change to value. Since from the invention of the internet in the early 1990s, the growth of the data has been increasing steadily. In Past Decade data generation growth is massively high which become a great challenge in storing, managing and process of data. This set a path to the new concept of Big Data, a concept that concerns with all generated data that are analyzed and processed in the day to day tools (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Jeong and Shin (2016) posted a security management scheme that allows users to easily access Big Data from different network environments. For implementing security management using key management, they added furthermore as future research as to Design and operate a model that can integrate and manage the stratified properties of the security awareness information sent and received between users and servers (Bakshi, 2012).

RELATED WORK

Literature Survey

Liu et al. (2014) proposed a mining system with Big Graph analysis by performing in bulk synchronous parallel (BSP) naming it as BSP based Graph Mining (BSPGM). This System inferred is compared with Hadoop Map-Reduce concept in processing Massive Data and it is developed based on Cloud platform. The drawback of this system often it restricts the graph data in the processing phase.

Meng et al. (2014) suggested keyword-Aware service recommendation method using Hadoop when customer service is growing rapidly with online information generation is difficult to use traditional service recommendation system for large scale data. This lacks scalability and inefficiency in processing the massive data. Usually recommending system uses same rating and ranking of service but here the author uses frequently used keyword for search of particular word and they are analysed. The author uses MapReduce framework.

Chen, Mao, and Liu (2014) put forth system as privacy aware cross cloud service for Big Data application using MapReduce. This paper uses medical data as dataset in large scale for analysis the data. They propose Hiresome-II (History record based service optimization method). Cloud uses QaS for history records. Complexity of the cloud system is resolved using Hiresome-II.

Jamshidi et al. (2015) discuss how Big Data and system of service (SoS) works together. Big Data analytics tools used in this paper are Principal Component analysis (PCA), fuzzy logic clustering and K means. These tools of Big Data help in handling (a) Extract information. (b) Build acknowledge from derived data. (c) Develop model for Big Data.

Elgendy and Elragal (2014) explain the Large data graph analyzing by Big Data. Graph processing is implemented using the shortest path of the data. Map Reduce is used to determine the graph processing.

Boyd and Crawford (2012) discuss the question raised by various different field of technology where huge data are generated, processed and handled. The author discusses the varies tools from which Big Data are processed and studied. Furthermore, author an intimate user who uses Big Data and its provocation about its mystery.

Related Content

IoT-Fog-Blockchain Framework: Opportunities and Challenges

Tanweer Alam (2020). *International Journal of Fog Computing (pp. 1-20).*

www.irma-international.org/article/iot-fog-blockchain-framework/266473

Power-Aware Virtualization: Dynamic Voltage Frequency Scaling Insights and Communication-Aware Request Stacking

Dhaarini K. N. Hathwar, Srinidhi R. Bharadwajand Syed Muzamil Basha (2024). *Computational Intelligence for Green Cloud Computing and Digital Waste Management (pp. 84-108).*

www.irma-international.org/chapter/power-aware-virtualization/340523

Predictive Modeling for Imbalanced Big Data in SAS Enterprise Miner and R

Son Nguyen, Alan Olinsky, John Quinnand Phyllis Schumacher (2018). *International Journal of Fog Computing (pp. 83-108).*

www.irma-international.org/article/predictive-modeling-for-imbalanced-big-data-in-sas-enterprise-miner-and-r/210567

Edge Cloud: The Future Technology for Internet of Things

Lucia Agnes Beena Thomas (2019). *Novel Practices and Trends in Grid and Cloud Computing (pp. 107-131).*

www.irma-international.org/chapter/edge-cloud/230635

Submesh Allocation in 3D Mesh Multicomputers Using Free Lists: A Corner-Boundary Approach with Complete Recognition Capability

Saad Bani-Mohammad, Ismail Ababnehand Motasem Al Smadi (2015). *Advanced Research on Cloud Computing Design and Applications (pp. 171-187).*

www.irma-international.org/chapter/submesh-allocation-in-3d-mesh-multicomputers-using-free-lists/138504