Chapter 2
New Framework Modeling for Big Data Analysis of the Future

Mirza Tanweer Ahmad Beig
SGT University, India

Varun Kashyap
SGT University, India

Megha Walia
SGT University, India

ABSTRACT

The area of big data analysis confronts several obstacles in its quest to derive useful insights from the ever-increasing amount and complexity of available data. To cope with the future volume, velocity, and diversity of data, new frameworks and models must be created. In this article, the authors offer a new framework for big data analysis that makes use of a variety of recently developed tools and techniques specifically designed to meet these demands. The three main pillars of our methodology are data acquisition, data processing, and data analysis. To ensure effective and continuous data collection from many sources, the authors make use of recent developments in data streaming and real-time data processing methods. This guarantees that the framework can process large amounts of data quickly enough to allow for timely analysis. The authors do tests using real-world, large-scale data sets to see how well this suggested framework performs in practice. When compared to conventional methods, the results show dramatic enhancements in terms of processing velocity, scalability, and precision. The authors also emphasize the framework's potential for integration with cutting-edge technologies like edge computing and internet of things (IoT) gadgets, as well as its flexibility to accommodate shifting data landscapes. Enhanced decision-making and insights in the age of big data are made possible by the integration of state-of-the-art technology and techniques, which allow for efficient data intake, scalable processing, and sophisticated analytics.

DOI: 10.4018/979-8-3693-1355-8.ch002

INTRODUCTION

Different data processing systems, each with its own set of features and capabilities, have emerged in recent years to meet the expanding need for Big Data analytics. Both batch processing systems, which deal with data while it is at rest, and stream processing systems, which deal with data while it is in motion, are among the developed data processing systems (Angeline et al., 2023). In today's Big Data age, however, the benefits to businesses will accrue more from a platform that can accommodate both processing paradigms (Chen & Zhang, 2019). For instance, monitoring stream-based apps in real-time might uncover suspicious or malicious behavior (Chen et al., 2014). Further categorization, correlation through offline analysis, and appropriate action may be conducted in response to the occurrence (Chen et al., 2019). Applications that make use of both real-time and offline analysis might benefit greatly from a modelling framework that allows for both batch and stream processing (Chunduri et al., 2023). The requirements of the future generation of Big Data Log Analysis may be met in large part by such a system. One of the most significant areas of Big Data analytics is the topic of this study (Elgendy et al., 2018).

Many recent studies have focused on solving the difficulties of log analysis in real-time. Security incidents in the IT infrastructure may now be managed with ease thanks to advancements in big data analytics (Fernández & Fernández, 2019). Research into log analytics has shown that both stream and batch log analytics may be very helpful to administrators in keeping track of security occurrences and taking preventative measures. There are often many stages to a major security breach (Francis & Sheeja, 2023). That's why it's so important to see the warning signs of impending catastrophic security crises as soon as possible (Gaayathri et al., 2023). Securing an organization's IT infrastructure is of the utmost importance. Without monitoring, analysis, and correlation of relevant traffic, it is impossible to detect security weaknesses in deployed Systems, Networks, and Application Servers (Gandomi & Haider, 2015). Keeping an organization's security posture in line with its security policy requires the collection and analysis of massive amounts of log data, both in real-time and for historical study (Haider et al., 2024).

The exponential need for log analytics may be met with the aid of established technologies like Big Data, Cloud Computing, and Kubernetes clustering, all of which provide a number of benefits that have not before been communicated (Hoque et al., 2019). Students in the field of technology have conducted an in-depth analysis of the magnitude and relevance of the most promising ongoing and future developments in the field of log analytics and have developed unique technical performance enhancements for these applications (Kannan et al., 2022). Understanding the impact of Big Data, Cloud computing, and Kubernetes clustering research breakthroughs on existing log analytics procedures and evaluating performance via rigorous analysis is the primary goal of this study (Li et al., 2020). This chapter delves into the origins of log analytics as well as its history, the significance of Big Data analytics, the application of Cloud Computing to log analytics, the impetus for exploring this topic further, research objectives, hypotheses, problem definition, and methodologies (Lohith & Bharatesh Cahkravarthi, 2015).

This study introduces a Cloud- and intranet-deployable, centralized heterogeneous log analysis system built on top of Apache Flume, Apache Kafka, the ELK Stack, Spark, the AWS Cloud, and Kubernetes (Lohith et al., 2015). By using the features offered by each of the aforementioned open-source services and platforms, the framework strives to deliver a platform that integrates stream and batch processing and provides real-time search and analytics (Lohith et al., 2023). After stream and batch processing are completed to satisfy the existing log analytic requirements, this document may serve as a reference for implementing Elasticsearch-based data analysis (Lu et al., 2020). When we compared the performance of the unique big data log analytic framework we built to that of popular commercial and open-source