Integration of Big Data Analytics Platform for E-Repository in University of Kelaniya

Priyashantha, H.A.I.

inosha@kln.ac.lk

Abstract

Big Data revolutionize the organizations and institutions, improving their efficiency and performance with discovering valuable insights within the data itself which contain the high social, economic, and scientific value. The power of advanced analytics reveal hidden patterns, unknown correlations, consumer trends, user preferences and other useful information; to make significantly better and faster decision, it has been spotlighted all around the academia and industry. Big data analytics uses three main processes while generating a real value from data: data collection, data analysis and the presentation of information. Data analyzing process is the bridge that transforms data into information with a real value. In general, big data consist with four attributes, known as the volume, variety, velocity, and value.

Elastic stack or the ELK Stack is acollection of three open-source productsElasticsearch, Logstashand Kibanawhichdelivers actionable insight from structured and unstructured data, using with DSpace, digital scholarly publication platform in the University.Logstashcollects parses and indexes data flowing from DSpace server to Elasticsearchaggregationto explore trends and patterns indata.Kibanaprovides visualization capabilities on top of the content indexed on an Elasticsearch cluster. In the University of Kelaniya, such system was not adopted for discovering and evaluate scholarly repository usage.As the main component of this system, free and open source ELK stack deployed on existing Ubuntu server. Lightweight data shipper "Beats" installed in scholarly repository server, which sends data to ELK server.This implementation efficiently geo-identifies who accessed what and when, and what are the most accessed domains. In addition, thisuntangled the identification of correlations using queries and searches, and acquired rich visualizations and near real-time monitoring dashboards on top of that.

Keywords: Big data, elastic stack, dataanalytics