Big data streaming platforms: A review

Kumar, Harish[a]; Soh, Ping Jack[b]; Ismail, Mohd Arfian[c]
[a] Department of Computer Science, College of Computer Science, King Khalid University, Abha, 61413, Saudi Arabia
[b] University of Oulu, Finland
[c] Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, Malaysia

ABSTRACT
Yesterday's "Big Data" is today's "data." As technology advances, new difficulties and new solutions emerge. In recent years, as a result of the development of Internet of Things (IoT) applications, the area of Data Mining has been confronted with the difficulty of analyzing and interpreting data streams in real time and at a high data throughput. This situation is known as the velocity element of big data. The rapid advancement of technology has come with an increased use of social media, computer networks, cloud computing, and the IoT. Experiments in the laboratory also generate a large quantity of data, which must be gathered, handled, and evaluated. This massive amount of data is referred to as "Big Data." Analysts have seen an upsurge in data including valuable and worthless elements. In extracting usable information, data warehouses struggle to keep up with the rising volume of data collected. This article provides an overview of big data architecture and platforms, tools for data stream processing, and examples of implementations. Streaming computing is the focus of our project, which is building a data stream management system to deliver large-scale, cost-effective big data services. Owing to this study, the feasibility of large-scale data processing for distributed, real-time computing is improved even when the systems are overwhelmed.