

Abstract: The development of web 2.0 and related technologies have led to an exponential increase in various types of user-generated content including textual and networked information. Finding meaningful nuggets of knowledge from such a big and diverse data has attracted a lot of attention. Hadoop and MapReduce as a well known distributed environment and computing framework have been widely and successfully deployed in many domains, particularly in the field of cybersecurity and network analysis. Many scalable machine-learning algorithms such as K-means clustering, association rule mining, collaborative filtering, topic modeling, and network analysis have been proposed and implemented in many open-source packages (e.g. Apache Mahout). In this tutorial, we plan to discuss basic concepts, widely used algorithms, and some real-world applications in cybersecurity using MapReduce.

Bio: KZ is an assistant professor in the department of Decision, Operations & Information Technologies at the University of Maryland, College park. His research focuses on applying scalable Machine Learning, Natural Language Processing, and Social Network Analysis techniques on big data problems in business and healthcare. He published many papers in conferences and journals. He serves program committees for many international conferences and he is an associate editor for electronic commerce research. For more information, please look at his website: <http://www.terpconnect.umd.edu/~kpzhang>.

