MapReduce Programming Model in Hadoop and Big Data Workflow Systems

Mahdi Ebrahimi
Computer Science

4/24/2017 at 04:15 pm
H350

Abstract

MapReduce is a programming model and software framework first developed by Google to facilitate and simplify the processing of vast amounts of data in parallel on large clusters of commodity hardware. Hadoop is an Apache Software Foundation project that importantly provides two things: a distributed filesystem called HDFS (Hadoop Distributed File System) and a framework and API for building and running MapReduce jobs. In this lecture, I will talk about the MapReduce in Hadoop, and then I will present the MapReduce workflow that I developed via our big data workflow system called DATAVIEW (www.dataview.org). DATAVIEW is a big data workflow system for managing data analysis processes. A big data workflow is the computerized modeling and automation of a process consisting of a set of computational tasks and their data interdependencies to process and analyze data of ever increasing in scale, complexity, and rate of acquisition.

Bio

Mahdi Ebrahimi is currently a PhD candidate in the Big Data Research Lab at Wayne State University under the supervision of Dr. Shiyong Lu. His main research interest is in the field of big data management, with the focus on big data placement and task mapping optimization for cloud-based workflows. His broader interests include big data, data science, cloud workflow security, cloud computing, Internet of Things (IoT), and multi-objective optimization. He has published several research articles in peer-reviewed international journal and conferences, including International Journal of Big Data (IJBD), IEEE International Conference on Big Data (IEEE BigData), IEEE International Conference on Big Data Computing Service and Application (IEEE BigDataService), IEEE International Congress on Big Data (IEEE BigData Congress), and others. He is a member of IEEE and ACM.

Faculty Host: Dr. James Abello