Hadoop, an open-source implementation of MapReduce, is a data management framework on large cluster of commodity machines to handle data-intensive applications. Task stragglers in MapReduce jobs dramatically impede job execution on massive datasets in cloud computing systems. This impedance is due to the uneven distribution of input data and computation load among cluster nodes, heterogeneous cluster nodes, data skew in reduce phase, resource contention situations, and network configurations. All these reasons may cause delay failure and the violation of job completion time. Replica placement in Hadoop distributed file system (HDFS) plays a significant role in data availability and the balanced utilization of clusters. The first contribution in this dissertation presents an innovative replica placement policy for HDFS. It can perfectly balance the data load among cluster nodes. The heterogeneity of cluster nodes exacerbates the issue of computational load balancing; therefore, in the second contribution another replica placement algorithm has been proposed for heterogeneous cluster environments. The timing of identifying the straggler map task is very important for straggler mitigation in data-intensive cloud computing. To mitigate the straggler map task, Present Progress and Feedback based Speculative Execution (PFSE) algorithm has been proposed in the third contribution. Straggler reduce task aggravates the violation of MapReduce job completion time. Straggler reduce task is typically the result of bad data partitioning during the reduce phase. In the fourth contribution, a new partitioning scheme, named balanced data clusters partitioner (BDCP), is proposed to mitigate straggler reduce tasks. BDCP can assist in straggler mitigation during reduce phase and minimize the job completion time in MapReduce jobs within data-intensive cloud applications.

Major: Computer Engineering

Educational Career:
Bachelor's of Computer Engineering, BS, 2002, Almostanseriyyah University
Master's of Computer Engineering, MS, 2006, University of Technology

Committee in Charge:
Mostafa Bassiouni, Chair, Department of Computer Science
Mingjie Lin, Department of Electrical and Computer Engineering
Qun Zhou, Department of Electrical and Computer Engineering
Rickard Ewetz, Department of Electrical and Computer Engineering
Ivan Garibay, Department of Industrial Engineering & Management Systems

Approved for distribution by Mostafa Bassiouni, Committee Chair, on October 23, 2019.

The public is welcome to attend.