



Interdisciplinary Data Sciences

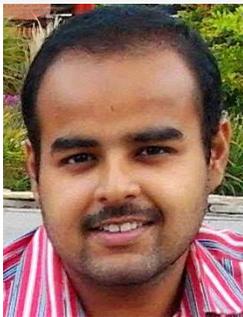


* IDSC Seminar Series *
October 20, 2017 10:00am-11:00am
Location: ENG 004
Presents

Dr. Shantanu Sharma
Department of Computer Science
University of California, Irvine

Title: Communication Aspects in MapReduce Computations

Abstract: MapReduce has proven to be one of the most useful paradigms in the revolution of distributed computing, where cloud services and cluster computing become the standard venue for computing. A MapReduce algorithm can be described by a *mapping schema*, which assigns inputs to a set of reducers, such that for each required output there exists a reducer that receives all the inputs that participate in the computation of this output. Reducers have a capacity, which limits the sets of inputs that they can be assigned. However, individual inputs may vary in terms of size. We consider, for the first time, mapping schemas where input sizes are part of the considerations and restrictions. One of the significant parameters to optimize in any MapReduce job is communication cost between the map and reduce phases. The communication cost can be optimized by minimizing the number of copies of inputs sent to the reducers. The communication cost is closely related to the number of reducers of constrained capacity that are used to accommodate appropriately the inputs, so that the requirement of how the inputs must meet in a reducer is satisfied. In this work, we consider a family of problems where it is required that each input meets with each other input in at least one reducer. We also consider a slightly different family of problems in which, each input of a list X , is required to meet each input of another list, Y , in at least one reducer. We prove that finding an optimal mapping schema for these families of problems is NP-hard, and present a bin-packing-based approximation algorithm for finding a near optimal mapping schema.



Biography: Shantanu Sharma received his PhD in Computer Science in 2016 from Ben-Gurion University, Israel, and Master of Technology (M.Tech.) degree in Computer Science from National Institute of Technology, Kurukshetra, India, in 2011. He was awarded a gold medal for the first position in his M.Tech. degree. Currently, he is pursuing his Post Doc at the University of California, Irvine, USA, assisted by Prof. Sharad Mehrotra. His research interests include designing models for MapReduce computations, data security, distributed algorithms, mobile computing, and wireless

IDSC Contact: **communication.**

Dr. K. Ramachandran
University of South Florida
4202 E Fowler Ave, CMC317
Tampa, FL 33620-5700
E-mail: ram@usf.edu
Telephone: (813)-974-1270
Fax: (813)-974-2700
<http://idsbigdata.com/>