

Performance Optimization, Elasticity and Sustainability in BigData Cloud Computing

Prof. Xiaobo Zhou

Collorado University

Wednesday, May 27, 2015

3:00 pm - 5:00 pm

Room: Ratssaal, Andreas-Pfitzmann-Bau

Abstract: MapReduce, a parallel distributed programming model on clusters of commodity hardware, has emerged as the de facto standard for big data analytics. Big data analytics usually require distributed computing at scale which is often hard to afford for small business and institutes. Large MapReduce clusters also suffer from issues like low cluster efficiency. Cloud computing, unlocked by virtualization technology, allows the creation of dynamic virtual clusters with elastic resource allocation. Moving MapReduce into the cloud seems to be the promising future of efficient and affordable big data analytics. However, the heterogeneity, high dynamics and multi-tenancy of the Cloud, and the semantic gap between the MapReduce runtime and the resource manager of the cloud platform make building elastic and efficient virtual MapReduce clusters very challenging. This talk focuses on performance optimization and energy efficiency of BigData processing in the cloud. Specifically, it presents Ant, an adaptive task tuning approach that aims to automatically find the optimal MapReduce parameter settings for individual tasks running on different nodes, be it in a physical heterogeneous cluster or a virtual cluster in the cloud. It then introduces FlexSlot, a flexible slot management for moving Hadoop into the Cloud. FlexSlot provides a user-transparent task slot management scheme that coordinates the slot-based resource management of Hadoop with underlying cloud infrastructure. Furthermore, this talk discusses energy efficiency and sustainability of datacenter cloud computing. In specific, it presents ePower, an elastic power-aware resource provisioning of heterogeneous workloads in self-sustainable datacenters. The talk will conclude with discussions for potential research collaborations in the related fields.

Bio: Dr. Xiaobo Zhou is a Professor and the Chair of the Department of Computer Science, University of Colorado, Colorado Springs. He leads the Distributed, Sustainable, and Cloud Computing Lab. His research lies broadly in computer network systems, more specifically, Cloud computing and Datacenters, BigData parallel and distributed processing, autonomic and sustainable computing, distributed systems and middleware, and scalable Internet services. His research has been continuously supported by the US National Science Foundation. He was a recipient of NSF CAREER Award 2009, and the University Faculty Award for Excellence in Research 2011. His work won the Best Paper Award of USENIX ICAC 2013 and a Best Paper Nominee of IEEE HPCA 2013. Dr. Zhou served as a Program Co-Chair of the IEEE/ACM CCGrid 2015 and IEEE ICCCN 2011, a General Co-Chair of the IEEE ICCCN 2014 and IEEE ICCCN 2012, a Program Vice Chair of IEEE/ACM CCGrid 2014, IEEE GLOBECOM 2010, ICCCN 2009, HPCC 2008, and IEEE/IFIP EUC 2008, and the Workshop General Chair of the IEEE ICCCN 2007 and IFIP EUC 2006. He served in several international journals including ACM Transactions on Autonomous and Adaptive Systems, the Journal of Parallel and Distributed Computing, and Computer Communications as a guest editor or associate editor. He obtained his BS, MS and PhD degrees in Computer Science from Nanjing University, China.

