Resource Management Problems in Cloud Computing

Richard T.B. Ma
National University of Singapore

Abstract

Through virtualization technologies, cloud computing enables users to easily acquire computation resources; however, the capacity of private clouds and the purchasing budgets for third-party resources are often limited. As the performance and quality of service (QoS) of software applications are directly influenced by the amount of resources they use, how to allocate the limited resources among users becomes a crucial decision for private cloud providers to make.

In this talk, we will explore designs of resource management frameworks that achieve various objectives. First, we will focus on a multi-user scenario and discuss Abacus, a generic resource management framework that interacts with users to allow them to specify their priorities and job characteristics, and allocate and schedule resources accordingly. Abacus provides efficient service differentiations based on an auction mechanism, which ensures desirable properties such as incentive compatibility and monotonicity. Second, we will consider data streaming applications with real-time constraints and discuss DRS, a dynamic resource scheduling and optimization framework. DRS measures instantaneous performance metrics and uses a performance model based on open queueing network to optimize resource allocation decisions. As efficient dynamic resource allocation needs system supports for elasticity, in the last part, we will discuss our on-going work of Elasticutor, an executor-centric design of streaming system that enables timely elasticity for real-time processing.

Biography

Richard T. B. Ma received the B.Sc. (first-class honors) degree in computer science in July 2002 and the M.Phil. degree in computer science and engineering in July 2004, both from the Chinese University of Hong Kong, and the Ph.D. degree in electrical engineering in May 2010 from Columbia University. During his Ph.D. study, he worked as a research intern at IBM T. J. Watson Research Center in New York and Telefonica Research in Barcelona. He is currently an Assistant Professor in School of Computing at National University of Singapore and a Research Scientist in Advanced Digital Science Center, University of Illinois. Richard is a co-recipient of the best paper award in the IEEE IC2E 2013, IEEE ICNP 2014 and IEEE Workshop on Smart Data Pricing (SDP) 2015. His current research interests include economics and evolution of the Internet, performance evaluation, big data analytics and cloud computing.