2012/2013 Seminar Series

www.cs.unb.ca/seminarseries

UNB Faculty of COMPUTER SCIENCE

Fredericton · New Brunswick · Canada

# On Data Staging Strategies for Mobile Accesses to Cloud Services


**By:**

**Dr. Yang Wang, Post-Doctoral Fellow, IBM Centre for Advanced Studies (CAS) Atlantic, University of New Brunswick**



With increasing data accessibility demands on Clouds, data availability maximization seems to be an important problem to consider to maintain high-fidelity and time-bounded service expectations in Clouds. For example, one of the pressing needs by the Cloud Service Providers (CSPs) is to efficiently serve the needs of the user requests that demand single or multiple data items in the shortest possible time. Thus with growing population of Cloud users, the problem of making the requested data available to the users becomes an imperative issue for CSPs to guarantee high quality services. An particularly appealing approach to maximizing such data availability is to stage the requested data to some vantage sites and cache the data for a period of time so that the quality of service (e.g., latency minimization, network traffic reduction) for user's future accesses can be greatly improved.

In this talk, I will introduce the strategies for efficiently achieving data staging and caching on a set of vantage sites in a Cloud system with a minimum cost. Unlike the traditional research, we do not intend to identify the staging time and locations to facilitate the future requests. Instead, with such a kind of information presumably known in advance, our goal is to stage the shared data items to pre-determined sites at pre-specified time instants in an efficient way while minimizing the monetary costs. In particular we follow the homogeneous cost model and fully network model and extend the analysis to multiple data items, each with single or multiple copies. Our results show that when the ratio of transmission cost rate and caching cost rate is low, a single copy of each data item can efficiently serve all the user requests made for any subset of the total distinct data items. When multiple copies are allowed to optimize the cost, we also consider the trade-off between the transmission cost and caching cost by controlling the upper bounds to the number of transmissions or copies. The upper bound can be either on per-item basis or on all-item basis. Optimal solutions in polynomial time to all these cases are also presented provided that the upper bound is polynomially bounded by the number of service requests and the number of distinct data items. We validate our findings by implementing a data staging solver, whereby conducting extensive simulation studies.

Wang Yang is currently working as a PDF with IBM Center for Advanced Studies (CAS), Atlantic, University of New Brunswick, Canada. Before joining CAS Atlantic in May, 2012, he was a research fellow at the National University of Singapore from Nov. 2010 to May 2012. In 2009, he was awarded by Alberta Ingenuity of Canada to be an Alberta industry R&D associate to conduct applied research in Madentec Ltd, Canada. Prior to that, he was a research associate in University of Alberta, Canada from Aug. 2008 to Mar. 2009. His research interest includes scientific workflow computations and virtualization technologies in Clouds and resource management algorithms. In addition he is also interested in exploiting heterogeneous multicore processors to accelerate the execution of the Java Virtual Machine. He received BS degree in Applied Mathematics from Ocean University of China, and MS and PhD (2008)


**Wednesday, October 3 @ 3:30pm**

**Information Technology Centre ITC317**