Hybrid Cloud and Cluster Computing Paradigms
For Scalable Data Intensive Applications

Dr. Judy Qiu
Computer Science Department
Indiana University

Date and Time: April 15, 2011 - 11:00am-12:00pm

Location: SEC 3437

Abstract:

Clouds and MapReduce have shown themselves to be a broadly useful approach to scientific computing especially for parallel data intensive applications. However, they have limited applicability to some areas such as data mining because MapReduce has poor performance on problems with an iterative structure present in the linear algebra that underlies much data analysis. Such problems can be run efficiently on clusters using MPI leading to a hybrid cloud and cluster environment. This motivates the design and implementation of an open source Iterative MapReduce system Twister. Comparisons of Amazon, Azure, and traditional Linux and Windows environments on common applications have shown encouraging performance and usability comparisons in several important non iterative cases. These are linked to MPI applications for final stages of the data analysis. Further we have released the open source Twister Iterative MapReduce and benchmarked it against basic MapReduce (Hadoop) and MPI in information retrieval and life sciences applications. The hybrid cloud (MapReduce) and cluster (MPI) approach offers an attractive production environment while Twister promises a uniform programming environment for many Life Sciences applications.

Biography:

Judy Qiu is an Assistant Professor of Computer Science in the School of Informatics and Computing and an assistant director of Digital Science Center at Indiana University. Her research interests are parallel and distributed systems, Cloud computing and high performance computing. Dr. Qiu leads the SALSA project (http://salsahpc.indiana.edu) involving both professional staff and PhD students from the IU School of Informatics and Computing. SALSA focuses on data-intensive computing at the intersection of Cloud and multicore technologies with an emphasis on life science applications using MapReduce and traditional parallel computing approaches. Qiu is also active in program service and supporting diversity in computing, which include serving as a Program Co-Chair of the 2nd IEEE International Conference of Cloud Computing Technology and Science 2010 and on editorial board of International Journal of Cloud Computing.