



# The Center for Autonomic Computing at Rutgers University

A National Science Foundation Industry/university Cooperative Research Center

<http://nscac.rutgers.edu>

## Scheduling and Management of Data Intensive Application Workflows in a Cloud

**Dr. Suraj Pandey**

The Commonwealth Scientific and Industrial Research Organisation (CSIRO),  
Australia

**Abstract:** Nowadays, scientific experiments are being conducted in collaboration with teams that are dispersed globally. They share data and utilize distributed resources for conducting experiments. As a result, scientific data are replicated and cached at distributed locations around the world. These data are part of application workflows, which are designed for reducing the complexity of executing and managing large applications on distributed computing environments. In order to execute these workflows in time and cost efficient manner, a workflow management system must take into account the presence of multiple data sources in addition to distributed computing resources provided by platforms such as Grids and Clouds. Building upon existing workflow management concepts, we present workflow scheduling algorithms specifically designed for managing data intensive applications in a Cloud. We experiment these algorithms using several real world applications, such as, image registrations for Functional Magnetic Resonance Imaging (fMRI), Gravitational wave search as workflow, distributed intrusion detection, and so forth, and describe our experiences when executing them using Amazon Cloud services.



**Bio:** Dr Suraj Pandey is a research fellow at The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia. Before joining CSIRO, he worked as a post-doctoral research fellow at the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, Australia. His PhD thesis focused on scheduling data intensive application workflows on Grid and Cloud computing environments. Currently, at CSIRO, Dr Pandey is working on projects related to: scalability of PaaS in Cloud computing, Business Process Management and prediction techniques.

**Wednesday, October 19, 2011, 2:00 PM**

**Place: Room 538, CoRE Bldg.  
Busch Campus • Piscataway, NJ**