

July 22: Ravi Madduri, Large-scale Research Data Management and Analysis Using Globus Services



SYNOPSIS:

In this talk, Mr. Madduri will describe Globus Genomics, a system that was developed for rapid analysis of large quantities of next-generation sequencing (NGS) genomic data. This system is notable for its high degree of end-to-end automation, which encompasses every stage of the data analysis pipeline from initial data access (from remote sequencing center or database, by the Globus file transfer system) to on-demand resource acquisition by a specialized elastic provisioner (on Amazon EC2); specification, configuration, and reuse of multi-step processing pipelines (via the Galaxy workflow system); creation of custom Amazon Machine Images; and efficient scheduling of these pipelines over many processors (via the HTCondor scheduler). The system allows biomedical researchers to perform rapid analysis of large NGS datasets using just a web browser in a fully automated manner, without software installation or a need for any local computing infrastructure.

[Session details...](#)

BIO:

Ravi Madduri is a senior researcher at the Computation Institute and project manager in the Mathematics and Computer Science Division at Argonne National Laboratory. Mr. Madduri is also an innovation fellow at the Polsky Center for Entrepreneurship and Innovation at the University of Chicago. He led several successful large projects in NSF, NIH and DoE. His research interests are in building sustainable, scalable services for science, reproducible research, large scale data management and analysis. He is one of three key contributors to the National Institutes of Health \$100M Cancer Biomedical Informatics Grid (caBIG), which links 60 NIH-funded cancer centers and clinical sites engaged in cancer research. For his efforts in project management, tool development, and collaboration, Mr. Madduri received several Outstanding Achievement Awards from NIH in recognition of his work on caBIG project management, tool development, and collaboration. Mr. Madduri is a lead architect on the scientific workflow design and implementation project under the caGrid toolkit. Mr. Madduri leads the Globus Genomics project (www.globus.org/genomics), which is widely used for genomics, proteomics, and other biomedical computations on Amazon cloud and other platforms. He also architected the Globus Galaxies platform that underpins Globus Genomics and several other cloud-based gateways realizing the vision of Science as a Service for creating, maintaining sustainable services for science. Mr. Madduri is co-PI on NHLBI-funded CardioVascular Grid project, big data analysis working group chair for the NIH BD2K BDDS center, co-PI for DoE funded Portal for Data and Analysis for Cosmological Simulation (PDACS) project and a senior member of NSF-funded Framework to Advance Climate, Economic, and Impact Investigations with Information Technology (FACE-IT).

SUMMARY:

Topic: Large-scale Research Data Management and Analysis Using Globus Services

Speaker: Ravi Madduri, Senior Researcher, Computation Institute, and Project Manager, Mathematics and Computer Science Division, Argonne National Laboratory, and Innovation Fellow, Polsky Center for Entrepreneurship and Innovation, University of Chicago

Date: Wednesday, July 22, 2015

Time: 11 AM – 12 PM ET

You are invited to listen to Mr. Madduri's presentation in Room 2W908 in the NCI Shady Grove Building on Medical Center Drive or via WebEx.

Presentation: [Download the presentation slides.](#)

About the NCI CBIIT Speaker Series:

The National Cancer Institute (NCI) Center for Biomedical Informatics and Information Technology (CBIIT) Speaker Series is a bi-weekly knowledge-sharing forum featuring both internal and external speakers on topics of interest to the biomedical informatics and research communities. For additional information, including past speaker series presentations, visit the [CBIIT Speaker Series page](#).

Individuals with disabilities who need reasonable accommodation to participate in this program should contact the Office of Space and Facilities Management (OSFM) at 240-276-5900 or the Federal TTY Relay number 1-800-877-8339.