

Research Computing Advisory Committee

Minutes November 7, 2011 (taken by Erik Deumens)

Present: Paul Avery, Reed Beaman, Richard Deason, Erik Deumens, Rolando Millian, Felix Liu (guest)

Discussion

Four agenda items were discussed (see <http://www.it.ufl.edu/governance/advisorycommittees/documents/RCAC-AgendaFall2011.pdf>). Felix Liu Associate Director for Research IT at the HSC was introduced on his first day in his new job to the committee by Richard Deason.

Report: Sunshine Grid project

This project is funded by a New Florida Cluster award to a collaboration of UF with FSU and USF, with a budget of \$200,000 at UF. The project is creating shared cyber infrastructure across universities in Florida to help researchers be more productive. The project supports three specific projects:

- The use of the Cryo-electron microscope at FSU and a similar unit at UF.
- The high-energy physics experiment CMS performed at the Large Hadron Collider at CERN requires data processing and sharing at UF and FSU.
- The ocean-atmosphere coupling research at FSU and USF.

The activities of the three institutions in the award (FSU, UF, USF) has expanded to include UCF and UM. Three HPC Summits have been held (Mar at UF, June at USF, Sep at UM) and the next one in Dec will be at UCF. The Sunshine State Education and Research Computing Alliance has been created to build and operate a cyber infrastructure for the state of Florida and all its Universities. SSERCA will have a booth at Supercomputing 2011, to be held Nov 14-18 in Seattle.

The current activity is focused on deploying shared file system infrastructure developed for CMS at UF to allow ocean-atmosphere researchers at FSU, USF and UM to share data.

Discussion: Cloud services

Researchers are looking at cloud services for CPU cycles and for storage of data with safety from data loss by disaster provided through redundancy. Richard Deason reported on negotiations with Teramark, a company offering commercial cloud services distinct from the more widely know Google and Amazon cloud services. The total cost of using such services is similar to providing them in house, however, the flexibility to scale to large capacity is a distinct advantage that is very expensive to provide in house.

The discussion offered that UF Research Computing should provide in the HPC Center expertise to help researchers on campus to use cloud services, such as those offered by Amazon and Google. For this, they may want to operate and run a small cluster for developing, configuring, and testing cloud applications and virtual machines.

For storing TeraByte-sized data sets, special services need to be developed since cloud services do not address the full diversity of need: Some research projects need TBs of storage in the form of many smaller GigaByte-sized files and these can be efficiently transferred and accessed in the cloud; researchers who manipulate files of TeraByte size need infrastructure more like what is being developed under the Sunshine Grid project. This topic will receive much attention from the DLC subcommittee.

Internet2 is partnering with Box.net to offer a cloud service for file storage and management. This service will start in pilot-mode in January 2012. It is complementary in that it provides tools for sharing data files and manipulating and viewing them. There is a limit on individual file size of 2 GB, but no limit on total amount of data stored. This service is somewhat similar to a nationwide SharePoint service.

The committee feels that it is of utmost importance that UF provides a coherent infrastructure to hold and manage research data to make sure its faculty members are competitive in the proposals. An example of an awarded, ongoing project is the IbiogIO project with Prof. J Fortes as one of the PIs. This project will collect images from museums at the rate of 1 PetaByte per year.

Discussion: Manage HIPAA, PHI, ITAR data

There is a need to support projects with large scale compute capability on data that is restricted. Currently, a number of ITAR projects are supported on the HPC resources. The HPC Center has put in place measures to secure data similar to what the DoD HPC sites have.

The committee requests that a list of requirements be compiled for all these categories of data so that we can align addressing them all. For example FISMA requires that 168 controls be in place. It is likely that several requirements are similar and can be addressed by a single set of security measures. Implementing a storage system that meets these requirements is important to make our faculty productive and competitive.

It is pointed out that teaching researchers how to use de-identified data, will alleviate the problem, as a large fraction of the research can be carried out effectively this way. For the remaining researcher projects, more strict procedures must be put in place.

Guest: Dr. Jianhua (Felix) Liu, Ph. D.

Dr. Liu is the new associate director of Research IT for UF&Shands Academic Health Center, starting on Monday November 7, 2011. He comes from the Ohio State University Medical Center and has experience in building IT infrastructure such as data warehousing for healthcare and medical research data. He is interested in coordinating his activities with the Research Computing Advisory Committee, for example his experience in working with researchers on using de-identified data can be of great value to the activities discussed in this meeting.

Next meeting will be of the RCAC DLC subcommittee on Nov 21 at 1:30 pm in NPB.