SRCPAC Minutes – Fall 2019
Date: Tuesday, December 3, 2019
Time: 10:00 a.m. – 11:30 a.m.
Location: 523 Butler Library
Meeting Called By: Chris Marianetti, Chair of SRCPAC

WELCOME & INTRODUCTIONS

Chris Marianetti called the meeting to order and asked that each person introduce themselves.

HIGH-PERFORMANCE COMPUTING UPDATE
Kyle Mandli, Chair of the HPC Operating Committee
George Garrett, Manager of Research Computing

George presented an update on HPC. He reviewed the purpose of the HPC Operating Committee for special requests related to using the cluster.

Support services were highlighted, referencing workshops throughout the semester, and working with Foundations for Research Computing for additional workshops (i.e. Amazon Web Services Cloud Computing, Intro to Singularity, Intro to Docker).

Cloud computing consulting is available to review features, cost estimates, and imaging or software needed to use cloud computing.

Terremoto cluster expansion is going live later this month (Dec. 2019), adding additional storage, nodes, and new research groups. There were over 14 million core hours utilized under Terremoto this year. The first set of Habanero cluster nodes expire in December 2020, with 222 nodes expiring, the remaining 80 will expire in December 2021. Habanero reached 41 million core hours utilized this year, totaling for 55 million core hours with Terremoto. It was discussed that this is cost-effective, as an estimate indicated it would cost $10M to do the same work through Amazon.

Spring 2020 Cluster is currently in planning stages. New machine type is likely, as the current model purchased under Terremoto will no longer be available. Purchasing would commence late Spring 2020, going live in late Fall 2020.

Other HPC Updates included a review of Singularity and Open OnDemand HPC Web Portal. The Data Center Cooling expansion was completed in July 2019, assuring capacity for HPC for the next several generations.

FOUNDATIONS FOR RESEARCH COMPUTING UPDATE
Marc Spiegelman, Chair of the Advisory Committee
Barbara Rockenbach, Associate University Librarian for Research and Learning

Marc and Barbara presented an update on Foundations for Research Computing, highlighting progress towards goals in the 2019/2020 Academic Year. There has been an increase in the number of introductory and more advanced programming in the form of bootcamps, intensives, workshops, and trainings. The Distinguished Lecture series continues with a series of four lectures over the course of the academic year, partnered with the Data Science Institute and Brown Institute.

The application process for the bootcamps, etc. has been revised so it provides a better pool of applicants by including a self-assessment which is then reviewed by the Foundations team. There is still a surplus of applications for this, so need to find a way to expand capacity to meet the demand. Foundations has also launched their Curricular Innovation Grant which allows graduate students and postdocs to create curriculum modules in 1-2 hour trainings for RC Foundations programming. The email went to over 2500 graduate students and postdocs, with 22 applications for 6 grants. The topics selected by the committee are:

1. Interactive Data Visualization with R & Shiny
2. Intro to Deep Learning with PyTorch
3. Wrangling Multilevel Data with R & the Tidyverse
4. Data Analysis and Manipulation with Xarray
5. Python for the Analysis and Visualization of Biological Datasets
With support from Quantitative Methods in the Social Sciences:
6. Tidying Survey Data in R
7. Data Visualization in R (ggplot2)

There was then a Q&A period where Marc and Barbara recorded questions and will be following-up with responses. Some of the questions/discussion include: Programming related to AR/VR, Interest/idea of parallel computing environments, More specialized course development, Using Foundations website as a resource aggregation source, Promoting or finding central MatLab or Python person.

RESEARCH DATA PLATFORM
Maneesha Aggarwal, AVP, Academic and Research Services

Maneesha presented on the new Research Data platform and reviewed it the purpose of it. The platform originated under a pilot with Columbia World Projects in an effort of "sharing data for the greater good", where data was to be collected for different projects, different types of data sets, and in different environments. The concept followed was that of the academic cloud, with connected datasets across disciplines, allowing for easy-access, providing options for data simulation, analysis, etc., and also keeping it secure, scalable, and sustainable.

The current model is set to only pay for the data you need. Within the platform, users can upload, store, or create data, share and discover leading to potential collaborations, analyze and visualize data, do web-based publications, and archive their data all within the platform.

OTHER BUSINESS

Chris opened the floor to discuss other business, particularly a discussion on CloudBank. Kyle Mandli mentioned that it is going well and some people have already begun moving over to it, with the current CloudBank running from 2019-2024. Columbia is having a workshop at UC-Boulder to discuss how to use CloudBank, and they are working at an open solution on how we could integrate with it.

CLOSING REMARKS

Chris Marianetti thanks everyone for their input and participation and closes the meeting.