Piloting Sustainable HPC for Research at Columbia

A position paper for the NSF Workshop on High Performance Computing Center Sustainability, May 3-5, 2010, Cornell University

Rajendra Bose, Ph.D., Manager, Research Computing Services
Alan Crosswell, Associate VP and Chief Technologist, Information Technology
Victoria Hamilton, Director, Research Initiatives, Office of the Executive VP for Research
Nilda Mesa, Assistant VP of Environmental Stewardship
Columbia University, New York NY
February 19, 2010

Introduction

Four factors are driving a re-examination of computational tools for research and education at Columbia University. One, trends such as increasingly sophisticated simulation tools, accelerating global collaboration and access to burgeoning datasets are driving interest in using computational tools broadly throughout disciplines—not just in science and engineering, but also in social sciences and the humanities. Two, many of these tools and programs require high performance computing (HPC) clusters that require a different level of infrastructure, power and cooling, and systems administration than yesterday's workstations. Three, at the same time, fewer of these disciplines have an intrinsic interest in the computational tools per se, so there is less expertise or educational value in managing the tools within a lab or department. Four, industry innovation in terms of energy efficient computers and virtualization software enable exploitation of economies-of-scale; this may enable wiser use of University resources than the current highly distributed model of multiple independent computing clusters managed by graduate students and postdoctoral researchers.

There is interest within some groups at Columbia for piloting sustainable, central HPC for targeted sets of researchers, and integrating these activities with the larger University goals of environmental stewardship. These plans are discussed in the following sections. We are also interested in partnering with others on a regional, statewide or national level to solve the sustainable HPC problem, and participation in the NSF Workshop on High Performance Computing Center Sustainability represents a promising start for this goal.

Planning for HPC Sustainability

In ongoing discussions about research computing at Columbia, principals involved with information technology (IT) and research support are focusing on the benefits of centralizing HPC purchasing, management and support specifically for small- to mid-sized (1-20 person) research groups. Our immediate concern is not the well-funded, technically savvy and technically interested research groups (sometimes called "Tier 1" groups) that are content to be independent of central IT. These Tier 1 groups are aligned with the post-mainframe-era academic tradition of leaving faculty to their own devices—and funding—to carry out their varied and successful research programs within their own labs or disciplinary communities.

Instead, we believe central HPC services will have the most impact on the growing number of groups below Tier 1 which must use HPC to stay current within their research discipline, or are interested in investigating HPC for application to their research for the first time. In both cases we are concerned with research groups which do not have the resources or interest to install, run and maintain their own, self-purchased HPC systems.

An internal University task force, in effect during 2008, interviewed over fifty departmental chairs and researchers on research computing topics. HPC was found to represent a priority area. Initial

reports from the task force reasoned that providing centrally administered, shared HPC resources to the growing number of faculty and graduate students across departments and schools requiring HPC would:

- facilitate research that would have otherwise required orders of magnitude more time or simply could not be done at all;
- bolster the ability to secure extramural funding for this type of research;
- return faculty and students to their primary responsibilities of research, teaching, and learning rather than spending time as system administrators for their own computing clusters;
- enable researchers to explore the value of HPC to their research and teaching without (or prior to) committing to the effort of raising funds to buy and establish a computing cluster;
- improve research computing security, stability and continuity through the involvement of professional IT teams and the use of standard practices for system maintenance and long-term planning;
- increase energy efficiency and decrease overall operating costs with central data center facilities for research computing; and
- promote sharing of computing resources which would help the University's Office of Environmental Stewardship meet the University's goals for green initiatives, including ambitious New York City and State greenhouse gas emission reduction targets.

Planning for sustainable HPC at Columbia is crucial for the careers of many faculty and staff researchers, and active discussion is taking place concerning the appropriate funding and purchasing models to use. These conversations have been fueled by the success of the first year of a three-year shared HPC pilot project, an on-going study of "green" practices in extant data centers funded by New York State, the possibility of NIH and other Federal grant funding for improvements to the University data center infrastructure, and other activities described below.

HPC and other research computing activities

A number of research computing-related activities are being pursued at Columbia:

- **Research computing group:** A four-person Research Computing Services (RCS) group was established in late 2008 to serve as Columbia's research community liaison with central Information Technology groups. RCS works with other related entities at Columbia, including the Office of Research Initiatives and the Libraries/Information Services, with the shared goal of advancing research at the University.
- **Shared HPC cluster:** Discussions during 2008 uncovered an opportunity to investigate the joint purchase and use of a centrally-managed computing cluster by Columbia researchers in two departments, Astronomy and Statistics, who were contemplating separate cluster hardware purchases. This 256 core HPC pilot system, housed in the University data center, was proposed at the end of November 2008, and has been operational since May 2009. This successful pilot introduced the paradigms of sharing the cost for computing resources and maintaining the active participation of researchers through a Governance Committee.

  RCS, working with other central IT divisions, is responsible for ongoing management, tests, and improvements of the HPC cluster, freeing academic researchers from these tasks. Research faculty hope to make this pilot system permanent, and other departments are interested in joining this shared system.
- **National and regional HPC outreach:** RCS is developing expertise in HPC topics through involvement with national initiatives, including the Teragrid Campus Champions program and the Open Science Grid. We also have contact with the New York State Energy Research and

Development Authority (NYSERDA) University Data Center focus, with the New York State HPC Program, developed by the Foundation for Science Technology and Innovation (NYSTAR), with New York University's HPC group, and with HPC groups in other regional institutions.

- **Embedded research computing support and consulting:** RCS serves as a Columbia researcher liaison for our pilot HPC system. The RCS staff also provides a successful example of "embedded" research computing support. One staff member is a central IT employee but reports to co-directors of a research center; he has an office at the research center, but reserves time to enhance his professional technical skills within RCS. Another staff member spends blocks of time as a research software consultant within a joint IT-Libraries social science data service.

In addition to the above activities, Columbia has pursued several grant proposals to request funds for improvements to its data center infrastructure. Columbia competed for and was awarded New York State Energy Research and Development Authority (NYSERDA) funds to roughly match the amount spent by the University for an *Advanced Concepts Data Center* demonstration project, begun in April, 2009. This project focuses on decreasing energy consumption and validating green data center practices in our existing 5,000 square foot central data center.

A proposal has also been submitted for funds to renovate infrastructure to provide a new *Core Research Computing Facility* that will consolidate computational resources and improve data storage options for over 25 NIH funded research groups at the University.

Central HPC and environmental stewardship

Rapidly-growing HPC research needs in the past years have resulted in an explosion of the number of computing clusters being deployed throughout Columbia. The University has also been heavily investing in expanding its multidisciplinary programs, constructing a new 14-story building which will mix scientists and engineers from different disciplines on each floor by research theme on the main campus, and developing a new 17-acre campus in the Manhattanville neighborhood of West Harlem. Columbia faces the challenge of how best to provide research HPC support for these new facilities.

In 2007, Columbia University became one of the original universities to join Mayor Michael Bloomberg's PlaNYC Challenge, which invited institutions of higher education in New York City to cut their greenhouse gas emissions by 30% in 10 years. As part of this commitment, the University's Office of Environmental Stewardship led an effort to conduct an internal assessment of greenhouse gas emissions and their sources. This baseline assessment found that over 90% of greenhouse gas emissions were attributable to buildings and how they function. It became apparent that one of the major sources of electricity demand in the non-residential buildings were computers and IT usage. At the same time that computing demands were being studied by our internal University task force, it became clear that interests were aligned: efficiencies both for IT and for cutting carbon emissions could be gained by identifying smaller computing clusters and duplication of capacity and moving towards a more streamlined system. In addition, the task force found that there were additional energy demands on building heating and cooling systems from computers and servers that were not located in areas with temperature controls or because the locations had not been designed with energy efficiency in mind. The climate action plan that has been developed to attain the greenhouse gas reduction goal includes a segment devoted to reductions from IT efficiencies.

We believe an effective approach to sustainable high performance computing is to integrate our ideas for central HPC support with the larger University goals of environmental stewardship. Our *Advanced Concepts Data Center* project and other recent grant proposals aim for a more efficient allocation of space and energy in the existing data center to host our HPC servers.