Workshop 1 Summary:

The key points that surfaced during workshop 1 are:
• There are many good reasons for small institutions not to make significant investments in hardware. We should take advantage of cloud and grid computing resources that already exist and are in the process of being developed, particularly those that are part of state infrastructure.
• Local support is necessary for faculty who wish to make use of cloud or grid computing resources.
• It is reasonable to have small clusters and specialized hardware at the local level, in which case system administration support is crucial.

Introduction:
Six upstate New York colleges (Bard, Colgate, Hamilton, Skidmore, Union, and Vassar) have received a one-year $40K grant from the Teagle Foundation in order to explore high performance computing in the liberal arts setting. Specifically, we are interested in identifying research opportunities that can make use of high performance computing capabilities (some of which may already be underway on our campuses) as well as determining the best way in which to make such capabilities available to faculty on our campuses.

The grant funds four workshops during 2009 with the goal of articulating current and near-term need for high performance computing at our institutions and suggesting preferred mechanisms for providing such facilities. In addition to workshop reports, we will generate a white paper that will be available outside our institutions so that other campuses can benefit from this effort as well.

The four workshops address the following issues:

1. What do we mean by "High Performance Computing"?
2. Applications of HPC to humanities and social science disciplines.
3. Applications of HPC to science and engineering disciplines.
4. How much access to HPC capabilities do we need and how should we get it?

Workshop 1 Overview:

Workshop 1, held on April 7, 2009, addressed the question of what we mean by "High Performance Computing". It included the following presentations (most are available at http://cs.union.edu/~barrv/Grants/Teagle/teagle-overview.html#W1):

• Geoffrey Fox, Indiana University, discussed the general landscape of HPC.
• Scott Kaplan, Amherst College, discussed the experience of setting up a large cluster and making it available to faculty across the campus.
• Peter Pacheco, University of San Francisco, discussed an undergraduate course in parallel computing.

• Brian Macherone, University at Albany and NYS Grid, discussed the NYS Grid and opportunities available for involvement by liberal arts colleges in the state.

• Micheal Ridley, NYSTAR, discussed the resources available state-wide and ways in which liberal arts colleges can tap into them.

Geoffrey Fox, Indiana University:

Fox presented the view that we are currently faced with a “data deluge” and we need to look at
• what hardware is necessary to analyze it
• what software applications are necessary to analyze it
• what digital library technology is necessary for dealing with data

Currently innovation is being driven by Web 2.0 technology and applications, not by enterprise software. Increasingly virtual organizations, supported by Web 2.0 technologies, are linking people, computers, and data. Data is naturally distributed, so distributed computing is easier than is parallel computing.

Fox compared grid computing to cloud computing. He argued that grids are still valid as a concept, but that lots of the technology involved with grids did not succeed because industry didn't adopt it. Industry likes clouds (cloud computing) instead of grids.
• Clouds are highly cost effective computing centers, with >100,000 nodes
• IBM and other large companies see cloud computing as a way of delivering computing to industry. Companies like Microsoft and Amazon are pouring huge resources into clouds.
• Clouds use virtualization
• Industry is not interested in parallel computing.

Clouds are the commercial solution to large scale computing, and “large scale computing” is, in his view, a better term than “high performance computing”. Clouds will replace grids, in part because they are based on commercial software. He sees commercial software as being more robust and maintainable than the academic software that is behind grid computing.

Is there any reason to have the computers located locally? May want to have specialized hardware, like graphics processing, located locally, and then have access to broad regional or national cyber infrastructure. Fox sees “cyber infrastructure” as including parallel and distributed computing capabilities, grid technologies, and Web 2.0 technologies.

Fox also talked about interdisciplinarity, building on his involvement in Informatics, which he defined as “interdisciplinary CS”. He said that for true interdisciplinarity we have to address the fact that the “other” field (the non-CS field) may have a very different view of how you do education. For example, Social Informatics and Bioinformatics are quite different because their roots differ – sociology and biology are very different in how they approach getting their students through their curricula.
Finally, Fox addressed the issue of multi-core technology. He said the technology is here today, and in 2-3 more years will be completely dominant. But the languages DARPA is working on will not be ready; they are working on a time scale that is too long. So there are lots of cores and nobody has really figured out what to do with all of them. He suggested that we should not try to teach the general programmer to program in parallel. We should work on parallelizing the libraries, like $\mathbb{R}$ and Matlab, so that when you feed data into those libraries it is processed properly. Then with students we can focus on how to do large scale simulations, address workflow issues, usability, tradeoffs, and the importance of coordination languages.

Scott Kaplan, Amherst College:

Amherst faculty from CS, physics, and geology received a grant that funded development of a cluster computer. The cluster was to be used by faculty across the math and science departments. Some faculty contributed text for the proposal, while others joined the project after the grant was received and construction of the cluster began. Ultimately the users were Amherst faculty, post-docs, outside collaborators, and students who were involved in thesis work and summer research.

They rejected requests to use just one or two nodes as fast computers. They are interested in use of the machine as a full cluster, not as a farm of individual fast servers.

Their experience taught them that cooling dictates capacity, that computational capability is determined by BTUs of heat that can be handled. The full scope of issues:

- Heat is a huge issue. Amherst has money left from the grant because they could not physically cool all the hardware they could afford to buy.
- System administration is a very serious job requiring someone who really knows what they are doing. Computers are cheap, people are expensive.
- Implementation support is critical. In order to really utilize a resource like the cluster, it is necessary that there be a capable support person, and the faculty have to be willing to put in time to get things up and running. There is little reuse within the college community, little amortization and generalization from project to project.

Kaplan presented some lessons learned:

- sharing was not a problem. As of the time of his presentation they had not yet experienced any resource overcommitment
- they had relatively few users and questioned whether the number of users justified the costs and the time investment. The cluster has become indispensable for some, and they expect the number of users will increase somewhat, but not a lot.
- There are no economies of scale for liberal arts colleges.
- Replacement is an issue. Now Amherst is committed to pay to replace the cluster when it ages.
- There are tradeoffs between using cloud computing and waiting for inexpensive machines to install locally. But even if they use grid or cloud computing, they would still need a person who would help faculty set up their problems. There are groups of non-expert faculty who would need implementation support.
Peter Pacheco, University of San Francisco:

Pacheco discussed efforts to integrate parallelism throughout the CS curriculum at USF. This was a response to the realization that most CS students are not learning how to work with the parallel hardware that is now available. Though students see concurrency in their OS course, typically very few actually take a course in parallel or distributed computing.

The solution at USF was to introduce a required lower level course in parallel computing, an upper level elective course in parallel computing, and support project-based courses and independent study for upper-level students. The lower-level course is a prerequisite for operating systems. Students take the lower-level course before they take algorithms, so they make performance determinations by writing code and actually timing it.

The remainder of Pacheco's talk focused on the content of the required lower level course, including information on hardware and software used and the classroom set up.

Brian Macherone, University at Albany and NYS Grid:

A 10GB LAN is being built to connect equipment and institutions across New York State. Working groups provide technical standards as well as middleware for connecting applications to the grid. They do not yet have any liberal arts colleges as participating institutions in the New York State grid.

Macherone issued an invitation for us to join the New York State grid which would allow us to share resources, take advantage of knowledge and technology exchange, as well as facilitate our connections to other communities within the grid.

In addition, Macherone issued an invitation from Christine Haile, CIO at UAlbany, for us to take advantage of UAlbany's expertise and facilities. In particular, they can

- help us get grid certificates and get onto the grid
- support us in updating our existing cluster computers
- consult with us on the design of new local cluster computers and help us bring them online.

Michael Ridley, NYSTAR:

Ridley discussed several NYSTAR programs that involve high performance computing. The HPC Allocation Program allocates resources to business, academic, and government institutions. The HPC Assistance Program assists in the use of HPC assets in New York State. Under this program, New York is hiring six computational scientists who will help people statewide. Multidisciplinary collaborations are of particular interest, especially if they can help address grand challenges. The computational scientists will assist for a period of two to three weeks, helping to get an application ready to run on state equipment. Priority will be given to projects that align with existing state investments (e.g. nanotechnology related projects).

There will be an HPC hub at UBuffalo, adding to the facilities at Stonybrook and RPI. Ridley made the point that the state is basically creating a cloud, creating a statewide
cyberinfrastructure resource that can then be shared throughout the state.

Ridley suggested that within our consortium we should focus on skills, tool use, and simulation. He strongly urged us not to invest in hardware. He thought it was reasonable that we have small clusters at our institutions, but that there was no point in each school making million dollar investments (plus the cost of power and support). He also urged us to take advantage of small state micro-grants (< $50K) to partner with researchers throughout the state on shared problems. This would give faculty at our small schools the opportunity to partner with researchers at the larger universities.

Finally, he encouraged us to utilize NYSERNET rather than paying for the commodity Internet. Both Macherone and Ridley raised the importance of ample network bandwidth at and within each institution in order to be able to utilize grid and cloud computing resources.

Closing Discussion:

During a closing discussion with all the speakers a number of final points and suggestions surfaced:

1. there is a real need for expertise in parallel computing areas
2. lots of local research is still being done on relatively small clusters
3. after the workshop series is over we might want to have a Computational Infrastructure day, run by NYSTAR, modeled on similar activities by the Teragrid, that would expose us to work being done and facilities that are available throughout the state.
4. We should all be on the NSF RSS feed.
5. Can we automate the toolset(s) so that it is easier for people to take advantage of HPC?