Plato Grant Proposal

The Advanced Computing Research Lab

Sherri Shulman, Neal Nelson, Isaac Overcast

Goals and Needs

The Advanced Computing Research Lab (ACRL) is a joint project initiated by advanced Computer Science students and Computer Science faculty. The ACRL itself is currently structured around a cluster of 20 computers. This means that although there are many different computers, they effectively function as one computer on the user end. This is what is known as distributed computing. In order to create our lab, we need a server powerful enough to task out jobs, monitor traffic, and offer enough extensibility to deal with any additional computers we receive. Our needs are well beyond those of a simple network; we need a highly capable server to handle all the tasks. The students and faculty involved in this project have two significant goals for the ACRL. First, to build a physical infrastructure and set of opportunities for advanced, hands on learning in computer science that can be integrated into the curriculum. Second, to build a relationship between computer science and other departments by giving CS students the opportunity to create tailored distributed computing software for specific research applications. Utilizing modern cluster computing methods, students and faculty will be able to expand their research in ways only available at much larger universities. Though the increased power of distributed computing offers many new opportunities for researchers, the prohibitive expense of development generally restricts distributed solutions to large and well-funded projects. We intend to bypass this problem by using students to build small, domain-specific programs that offer direct solutions to posed problems. Over time, the collection of these software components will form the nucleus of a diverse library of functions that can be extended to other similar domains. By establishing our lab as a site for such collaborative work we hope to redefine Computer Science at Evergreen as not just a tool to other sciences, but as an independent field of research that can both advance our work in computer science and provide meaningful support to other scientific study at Evergreen.

Expected Project Impact

We expect that this project will affect the students working in the lab by giving them opportunities to work in an advanced computing environment. This has two aspects. First development of the lab itself (both the hardware configuration and the software) are opportunities to develop and use state of the art architectures. Second, exploring the potentials of these architectures for actual research problems in other domains will allow the students to apply their theoretical understanding of Computer Science to real applications. Because Computer Science requires a significant amount of skill and concept acquisition before a Computer Scientist can apply his or her abilities to other domains, it is often hard to foster these opportunities at Evergreen. In creating this lab, we will be able to support a physical distributed computing environment that will last beyond the current year or two. As important, we will be creating on ongoing set of skills and experiences on which we can build from year to year. Currently there are 6 advanced Computer Science students working in the lab, two faculty (Neal Nelson and Sherri Shulman), as well as the endorsement of Isaac Overcast in Academic Computing.

This grant is concerned with completing the construction of the hardware architecture: acquiring a more powerful server will make it possible for our cluster to perform at the required effective speeds. With the support of Academic Computing and the oversight of the faculty, the students are doing all of the work in setting up the cluster computing environment. If this phase of the development of the lab is successful, we will have a state of the art cluster computing environment that can be used to achieve our long term goals of supporting and collaborating with other scientific research at Evergreen. In the near term, we are planning to identify one or two research areas on which to focus for our first phase of development. With a working cluster computer environment we anticipate that future students and programs will be able to continue to use the lab effectively:

1. Advanced Computer Science students can join the lab and continue work on creating software components utilizing the distributed computing environment in collaboration with other scientific researchers at Evergreen.
2. As we gain more expertise in the application of this architecture to different kinds of research problems, we will be able to expand our reach and apply our tools either to related problems or to new problems.
3. The sciences at Evergreen should also benefit in that they will have access to a state of the art computing environment and the support of a dedicated group of advanced Computer Science students whose goal is to find innovative solutions to data and/or compute intensive problems.

Enhancing Evergreen's Curricula and Educational Innovation

As indicated above, the ACRL will enhance Evergreen's Computer Science curriculum in two aspects: one in the construction of the hardware and software architectures required to construct the lab and another in the construction of the software components we develop to support the problem domains we identify. In addition, because the goal of the ACRL is to provide an ongoing lab, we also are focusing on how to pass on the expertise that we develop from year to year, from one student generation to the next.

In addition, the ACRL has the potential to add to the curriculum in other scientific domains at Evergreen. With an accessible distributed computing environment and a cadre of advanced Computer Science students dedicated to producing custom software support, we hope other areas and programs at Evergreen will be able to consider more or different kinds of questions and experiments to explore.

Lastly, this represents an opportunity to do real collaboration between Computer Science students and other domains at Evergreen. This is valuable in several ways: Computer Science students have the opportunity to use their skills on real problems; it fosters collaboration between different areas; we can advance our own Computer Science knowledge of how to apply distributed algorithms as well as help our "clients" solve their research questions.

This is an area of Computer Science that is not well-defined. Although we know how to build a cluster computing environment, and we have access to software architectures and libraries that support computing in such an environment, the application of such collection to diverse domains and research questions is not an off-the-shelf commodity. We are excited to think that we can create a lab that will allow us to explore how these pieces can all work together. We are also excited to think that we can create an ongoing presence at Evergreen that will support these kinds of collaborative scientific explorations: this has not been the case up to now.

Computing Needs

Because we are setting up a cluster-based lab we need a single server to service the data requests that will be made over the network. We need hardware that will give us the level of memory, processor power, and necessary disk space to back up three separate client architectures in a 20-50 node network. None of the machines that we currently have are powerful enough to make a single server architecture feasible. Due to the current restrictions, we will be unable to proceed with all the facets of this project without an adequately powerful server.

Plans and Skills and Requirements for Success

There are three major pieces to the lab: the physical plant, the software architecture, and the development of the software components to directly support selected research problems. The first two pieces are already under development. The students working in the lab are all advanced Computer Science students and the work is well within their capabilities. The original idea for the lab grew out of the work the students were doing for their independent project in Student Originated Software (SOS). The faculty of SOS (Neal Nelson and Sherri Shulman) are both actively involved in supporting the students. In addition, Academic Computing is also supporting the lab and we have recently approved a spring internship for the actual lab support.

Looking to the future, we need to develop expertise in two areas:

1. Understanding how to use the cluster effectively
2. Understanding what kinds of problems will be best solved in such an environment.

We are working on the development of a test suite to acquire the kind of expertise we will need in order to be successful in evaluating the strengths and capabilities of our cluster. In addition we are developing a questionnaire and a rubric to allow us to evaluate potential applications. Once we have these in place, we will interview possible research candidates at Evergreen to select a proof of concept "client". These activities are all being conducted under the supervision of Sherri Shulman and Neal Nelson.

Assuming that we have a server of sufficient power, we believe we have all the skills and support to successfully complete a proof of concept.

Costs and Needs

Currently we don't have a server that is capable of handling the amount of information that we plan to process. Each of our donated machines has 1 gigabyte of ram; the server we are considering has 24. Cluster computing is used for massive data trafficking: it is necessary to have a machine powerful enough to process multiple and rapid requests. It will come with a 3 year warranty. This purchase is intended to be a one-time cost and maintenance of the server will be performed by our internship.

PowerEdge R300 Quad Code Intel Xeon X3363, 2.83GHz, 2x6Mb Cache, 1333MHz FSB, No Operating System: $2863.00

This includes support from Dell, as well as all components necessary to get the server running. Incidental costs (cables, etc) will be recycled from Academic Computing and donations from St. Peter's. There will be no software cost, and the only long term cost will come from power usage and cooling. Maintenance will be carried out by the student intern administrator.

Sharing Our Work

There are two aspects for sharing our work. First we want to share our work and experience with the next generation of students so that the lab continues to be a place where advanced Computer Science students can do effective work. To this end we are working on manuals, tutorials, and learning modules that will embody the results of our work and allow students to build up their own expertise to build on ours. We are also maintaining a wiki with relevant links and information. The second aspect is how to share our results. We plan on writing regular technical reports for each of the phases. We hope that we will be able to publish some of these, both in Computer Science journals reporting on our experiences using cluster computing in this form and in journals detailing the results of the research investigations we support.