Experiences with sustainability at the UF HPC Center

Erik Deumens and Charles Taylor
High Performance Computing Center
University of Florida
Gainesville, FL 32611
March 12, 2010

The HPC center at UF
The University of Florida High Performance Computing Center was started in 2003 as a faculty driven effort to enhance the HPC environment for researchers using HPC on campus. One goal was to join together and coordinate the efforts of many. A strategic plan was developed that called for three phases of growth and expansion. The HPC Committee acts as a board of directors and developed a sustainability plan. In the first half of 2009 the phase III expansion came online with a new cluster in a newly renovated machine room bringing the support scope to the entire University. The HPC Center now operates close to 3,000 cores, 130 TB of disk storage with e Lustre file system over InfiniBand. It has three full time staff members and a part time director and an annual expense budget of $10,000. More details can be found on the web site http://www.hpc.ufl.edu; the documents describing the policies of the HPC Center can be found at http://www.hpc.ufl.edu/index.php?body=about.

Sustainability issues
The main challenge throughout its 6 years of operation has been funding and sustainability. We had to ask for funding for the staff on a yearly basis. Space, power and cooling infrastructure was negotiated on a case by case basis. The cost for hardware, nodes, RAM, networks switches, storage systems, was borne by grant funds from faculty members who invested in the HPC Center with matching from Colleges and from the Office of IT. Now we have, for the first time, a three year agreement with the Office of IT and the major colleges on campus to fund the staff for three years starting July 1, 2009. The University is building a new machine room facility with 10,000 square feet of machine room space, half for enterprise computing and the other half for HPC. To get sufficient funds in a timely manner to meet demand and replace aging equipment is still a challenge. The appendix contains our latest attempt to define an attractive set of services and agreements that will convince faculty members to join efforts with the HPC Center.

Next we give a brief list of some of the issues that we encountered in the few years of operation and some of the opportunities that we found to be useful to build a case.

1. Different researchers have different reasons for joining or not joining the campus-wide HPC effort; the effort is worthwhile as long as you have a significant number of people willing to work on the effort. It is not necessary that everyone participates, nor is it advisable to force everyone to participate.
2. The current direction of technology evolution favors centralization again:
   a. The high density of cores requires high power and high cooling capacity and it is economically favorable to put all such equipment in one special structure, be it a building or a container park.
   b. Renovating rooms in existing buildings is very expensive.
   c. Maintaining machines on a building shared with people makes energy management harder, such as turning off AC and heat during the night and weekend when the people are not in the offices and classrooms.
   d. The complexity of modern clusters, with parallel file systems, support for MPI, fast interconnect, batch queue system, security management, a large variety of applications, to name some details, makes that system administration is a profession of its own. It is not a good use of human resources to make a graduate student be a system administrator instead of a research assistant.

3. A shared resource has benefits that are hard to quantify in terms of dollars, but are nevertheless real and substantial.
   a. A research group may be able to buy a 32 CPU cluster, but its use case calls for running thousands of jobs every few weeks. Investing the funds into a shared facility with 3,000 cores will allow them to submit thousands of jobs one evening and have the results ready in the morning, whereas the dedicated 32 CPU cluster would have to work on the jobs for days or weeks.
   b. One research group invests their funds in CPUs and RAM. However, one exceptional project needs a large storage system for a few days. If they purchased an isolated cluster, it would not make sense to include a large storage system, and reduce the number of CPUs, for the eventuality of this exceptional project. In a shared infrastructure, another investor, for whom the storage is essential, may allow work to proceed. Because the infrastructure is shared, the technical issues of allowing the project to proceed are already taken care of.

4. We found that it is essential to create an atmosphere of trust for the shared facility to be successful. Some researchers are concerned to give up full control. Even in such cases, it may be more cost effective for the institution to support the researcher in a shared environment as outlined in the services document included as the appendix.

5. Funding from the university in some form or other is essential to make the shared effort sustainable. This funding can be used as leverage to get more researchers to join the effort. For the university this funding provides a good return on investment because of the economy of scale: Shared infrastructure funds go further by supporting many research projects with fewer buildings and fewer staff.

6. A new need for support is emerging: As more people are using HPC, we find that more of them do not have the interest, not the time, or the skills to create the software that uses the hardware in an optimal way for their project. It becomes economically feasible to attach special staff to the HPC center with software development and science and engineering expertise to support researchers in projects with the duration of a few months. Something that individual research projects cannot afford.
Experience in Sustainability

Appendix: UF HPC Center proposal for services

UF HPC Center

Mission
The University of Florida High-Performance Computing Center is a faculty-directed facility with the single mission of providing high-performance computing resources and support to the faculty whose research depends on large-scale computing.

Oversight
The HPC Center is governed by the policies and procedures approved and adopted by the HPC Committee (link to HPC Committee information - chair, members, etc – need to add to website).

Funding
The UF administration and colleges have been generous in supporting the HPC Center and have provided much of the current infrastructure (space, cooling, personnel). The Center was founded under the premise that its cost should be shared among the faculty who use it and most of the current equipment (hosts, storage, networking) has been purchased with faculty research funds. This arrangement enhances the purchasing power of faculty while ensuring that they are vested in the success of the HPC Center. Therefore, it is extremely important that faculty who intend to use the center budget funds for it in their research proposals.

The use of professionally managed computing resources where the institution shares some of the cost is usually viewed positively by funding agencies. The HPC Center will gladly assist with those aspects of your proposal which pertain to computing budgets and resources.

Service Offerings

Support
Software and hardware support is an included service with every offering. The HPC Center exists to facilitate the use of HPC resources within the University's research community and includes a professional staff - two with Ph.D's in scientific fields – with years of high-performance computing experience including application tuning and development.

“We are here to help you, your post-docs, and students so that you can focus on your research.”

Shared-Resource Investor
The shared-resource investor contributes a fixed amount for the use of the HPC Center resources. The duration of this investment is three years and is based on the warranty period of the equipment purchased. In exchange for this investment, the investor receives a fixed NCU\(^1\) allocation determined by the investment amount and the current per-NCU equipment cost. The per-NCU equipment cost is set by the HPC Committee based on the most recent major equipment purchase. The current rate is $700 per NCU. The HPC Center does not tie the investment to any particular hardware.

The shared-resource investor has a first-claim\(^2\) priority to the investment allocation, as well as to any unused resources on the shared portion of the cluster. This gives the shared-resource investor short-term access to resources beyond the limits of his or her investment allocation when the resources are available. Shared-Resource is the service offering that is recommended for groups who consume a predictable level of resources on an ongoing

\(^1\) An NCU (Normalized Computing Unit) is defined as a single core with the necessary components to make it fully functional such as memory, storage, and networking.

\(^2\) A priority level which exceeds that of non-investors such that a non-investor job will never be scheduled ahead of an investor job when submitted under equivalent conditions. Note that the base priority is one component of a job’s overall priority which may be greater than or less than the base priority depending on several factors including the resources requested and the number of CPUs already in use

UF HCP Center 3 3/12/2010
basis.

**Trial Investor**
The trial investor may use the HPC Center resources as a shared-resource investor for a fixed period of time (one month) in order to assess the operational aspects of the HPC Center. For the purposes of the trial, the trial investor will be given an NCU allocation commensurate with the anticipated investment level. At the end of the trial period, the trial investor can either invest via one of the standard service offerings or continue to use the facility as a non-investor.

**Billed Investor**
The billed investor agrees to be billed for actual usage at a rate determined by the HPC Center and approved by the HPC Committee. This rate is currently $0.06 per CPU hour. Additionally, any storage used over a nominal base allocation will also result in a charge to the investor. This rate is currently $14.33 per TB-month.

The billed investor will have a base priority equivalent to that of the shared-resource investor. An important distinction is that billed investors will be charged for all NCUs consumed whereas shared-resource investors pay in advance and will not be billed for additional NCUs if they manage to exceed their investment allocation. This is the service offering that is recommended for researchers with frequent but unpredictable usage patterns and who are not able, or do not want, to commit funds in advance.

**Short-Term Investor**
The short-term investor is one who needs dedicated access to resources for a fixed period of time which is substantially less than the three-year time frame of a shared-resource investor. In this case a reservation will be made for the necessary resources and the investor will pay a fee equal to the prorated cost that a shared-resource investor would pay for the same resources plus an "uplift" to compensate for the loss of the shared resources in the global pool. The current rate for short-term investments is $25 per NCU-Month.

**Hosted Investor**
The hosted investor will have access to dedicated (non-shared) resources, purchased by the investor, but managed by the HPC Center. These dedicated resources will be configured and managed as all other resources within the HPC Center with the following important exceptions.

1. Only those users specified by the investor will have access to the hosted resources,
2. Unused cycles will not be available to investors in other service offerings. That is to say, the resources will be dedicated exclusively to the hosted investor.
3. The HPC Center will be compensated for managing the resources and providing hardware and software support in one of the following two ways as chosen by the investor.
   a) The HPC Center charges the hosted investor 10% of the cost of the cluster per year.
   b) After three years, when the manufacturer warranty has expired, the cluster becomes part of the HPC Center shared resources and the dedicated use by the users in the research group of the hosted investor comes to an end. This provides the HPC Center with about one year of use of the cluster as contribution from the hosted investor for hosting, managing and supporting the dedicated cluster for three years.

It is important to note that the hosted equipment will be configured with the same software environment as the shared-resource cluster and will be under the same management and maintenance policies. Root access will not be provided and customized operating environments (beyond application installations) are not included as part of the service. In short, the hosted investor's resources will only differ from other HPC resources in regard to user access. We will work with the hosted investor to purchase a hardware configuration to meet any specific requirements such as storage, memory per core, or interconnect.

**Non-investor**
The non-investor contributes no funds to the HPC Center and will not be billed for use of the facilities. The non-investor has the lowest base priority available and his or her jobs will run only if there are no investors competing for the same (shared) resources.
Experience in Sustainability

Preemption Queue
Finally, investor at any level may submit jobs to the preemption queue. Such jobs will have a priority above that of the non-investor and below that of an investor. However, these jobs will be preempted, that is to say, killed, if the resources they are using are needed for a subsequently submitted investor job. This queue is best used for short jobs or jobs that are easily restarted.

Examples

Example 1: Shared-Resource Investor
A group has computing projects which will require an average of approximately 50,000 CPU hours per month continuously. What should the HPC Center investment be?

There are 730 hours in an average month. Therefore, 50,000 hours corresponds to the continuous (24x7) use of 68.5 processor cores. The Phase III cluster investment rate is $700 per core so the suggested investment level for this research group would be about $51,000. Over three years of use that comes to $0.03 per CPU hour and includes space, cooling, network connectivity, storage and software support. The rate of three cents per CPU hour is a very low rate and is made possible because, as noted previously, the UF administration and colleges are paying for much of the infrastructure associated with owning and operating HPC systems. Thus each dollar of faculty funding goes almost entirely toward the equipment they will use.

Example 2: Billed Investor
Group UNKN needs to run calculations periodically. It is difficult to predict when it will need to perform calculations, but when it does, its members prefer not to wait in the queue any longer than necessary. The work submitted needs to be finished promptly and funds to pay for the time and storage used are available.

Group UNKN became a billed investor and used 14,686 CPU hours and 1.6 TB of storage in their first month. During the first week of the following month, group ABC was charged $881.16 for their CPU use and $22.92 for storage use. The charges are computed at the rates of $0.06/CPU-hour and $14.33/TB-month.

Example 3: Short-Term Investor
Group WKRP needs access to sixty-four CPUs for three to four months but does not need the resources beyond that time frame and does not believe they will be involved in similar projects in the future.

The HPC Center creates a 64-CPU reservation for the exclusive use of Group WKRP for four months. The cost to group WKRP is the prorated cost of $900 per NCU for three years or $25 per NCU per month.

$900 / 36 months * 64 CPUS * 4 months = $6400.

Example 4: Hosted Investor
Group LMNO must have immediate and dedicated access to compute resources and does not consider a shared resource a viable option. They prefer not to spend the time and effort required to acquire and maintain their own cluster. The department chair has indicated that there is no acceptable space to house a cluster large enough for their needs and the time to acquire the space would put the group months behind in their research.

Group LMNO becomes a hosted investor. The HPC Center designs and specifies a cluster that both meets the group's requirements and falls within their budget. The HPC Center staff then oversees the purchase, installation and configuration of the cluster. The HPC Center staff also configures a queue devoted to the group and restricts the cluster to running jobs from only that queue. No other jobs ever run on the group's cluster even if its CPUs are idle.

Group LMNO chooses to support the HPC Center for managing the dedicated cluster by relinquishing it to the HPC Center after three years at which time it becomes a shared resource. Group LMNO plans to buy a new dedicated cluster at that time.

Example 5: Non-investor
Group EXPM occasionally runs calculations to support their experimental work. They typically run a handful of
calculations each month requiring a nominal number of NCUs. The impact of their jobs on the cluster's load is small and they are welcome to do their infrequent calculations on the HPC Center facilities. They can run a limited number of jobs in the regular queue at low priority. They understand that they can also submit jobs to the preemption queue. Such jobs may be preempted by investor jobs and will need to be resubmitted when they are.

### Summary of Investment Option Costs

<table>
<thead>
<tr>
<th>Option</th>
<th>Cost Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shared-Resource Investor</td>
<td>$700 per NCU (one time, 3 year duration)</td>
</tr>
<tr>
<td>Billed Investor</td>
<td>$0.06/NCU-Hour, $14.33/TB-Month</td>
</tr>
<tr>
<td>Short Term Investor</td>
<td>$25 per NCU-Month</td>
</tr>
<tr>
<td>Hosted Investor</td>
<td>10% of equipment costs per year, divestment after three years</td>
</tr>
</tbody>
</table>