Observations and Recommendations **2/9/15**
Regarding NSF's Support for High-Performance Computing

Nancy Wilkins-Diehr, Wayne Pfeiffer, and Richard L. Moore, SDSC

Observations

1. NSF has invested in resources for high-performance computing for 30 years with several funding models that have varying strengths and weaknesses.

These models successively evolved from

- a few centers of comparable size, which procured and supported their own hardware;
- two PACI partnerships, which featured lead sites with large systems and a few partner sites with smaller systems;
- a single TeraGrid, where nearly-identical systems were installed at four sites; and
- the current Track 2/XSEDE model, where diverse hardware is annually competed and distributed among several sites and operational support is a hybrid of minimal site support and shared services provided by XSEDE.

In addition, two special competitions were held to procure leadership-class systems:

- TCS-1:LeMieux, which was fully installed at PSC in 2001; and
- Track 1:Blue Waters, which was fully installed at NCSA in 2013.

There are many lessons to be learned from these models that a longer paper might describe. Perceived strengths and weaknesses of each model are highlighted here.

The individual center model had many strengths: each center could procure systems that performed well for users rather than systems that "competed best" (e.g., on LINPACK); a full-featured environment, including archival storage was provided; and more sustained funding assured a cadre of capable staff. The primary weakness was the perception that the centers were not taking best advantage of expertise within the academic community.

The PACI model "spread the wealth" beyond the handful of original centers. The strengths were shared expertise among institutions providing resources and collaborations between computer and computational scientists that led to significant advances in domain science. Notable weaknesses were diffusion of funding that led to some systems not much larger than those at many research universities and increased management overhead.

The TeraGrid initially installed IA-64 systems of varying sizes at four sites with a high-speed network that connected them. The primary strength, not envisioned at the outset, was the emergence of science gateways. There were several weaknesses, however, including a management structure that discouraged adoption of alternate architectures; an expensive, under-utilized network; and the perception that lessened competition might not have motivated vendors to provide the best deals.

The current Track 2/XSEDE model was designed to introduce more competition to obtain higher performing resources (at least as measured by LINPACK) at lower capital and operating costs. This is the model's primary strength. However, there are numerous weaknesses: the competitions have not been held on a regular schedule with consistent funding; one competition failed because of an over-aggressive bid; some competitions have funded smaller systems with inadequate operational support based on the assumption that XSEDE will pick up the slack; and the instability from competitions has reduced staff morale and led to attrition and nomadic behavior at virtually all sites. Overall, this has led to a portfolio of resources that poorly match user demand.

2. NSF's annual investment in HPC resources has substantially declined with time in terms of constant dollars despite a growing demand from an expanding user base.

While the need for supercomputing has increased dramatically over 30 years, NSF's investments in computing have not kept pace. We use two systems at SDSC as an example.

In 1985, we procured a Cray X-MP/48 and ancillary equipment with an NSF hardware budget of $6M per year for the X-MP's four-year life, and we received an additional $6M per year for operations. The X-MP was serial number 6 of the fastest supercomputer that could be bought at the time. In 2015, we expect to receive $12M as a hardware budget for Comet ($3M per year averaged over its four-year life) and $3M per year for operations; Comet will be the second-largest system in the XSEDE repertoire. Even including ~$2M per year in operational costs borne by XSEDE shared services, today's funding is 2/3 the level of 30 years ago in absolute dollars and only 30% in constant dollars when an escalation factor of 2.20 from 1985 to now is taken into account (http://data.bls.gov/cgi-bin/cpicalc.pl).

While it is true that hardware costs per computing cycle have come down dramatically over 30 years, there has also been an incredible growth in the need for analysis. Many researchers would have had a hard time envisioning the need for a supercomputer in 1985. Today, computing is ubiquitous. Science is increasingly digital; data from all sources continues to increase exponentially; and the need for supercomputers in all fields is greater than ever. Science gateways have brought even more users to supercomputing, outpacing command-line users on both XSEDE and NERSC systems in 2014. Moreover, the number of publications per compute cycle can be exceptionally high for gateways. In 2014, 4,600 users of the CIPRES gateway published 478 papers while using just 12M core hours.

Recently we estimated NSF's investment in resources (excluding Blue Waters) by analyzing NSF awards since the Track 2 era began in 2006. Figure 1 shows the annual capital investments and the aggregate operations costs across systems, as well as 3-year averages for capital costs and combined capital/operations. The healthy ~$30M/year investment in capital originally planned in the first four years was diminished by the unsuccessful third procurement (because of vendor default with no replacement). Subsequent to that, the capital investments have been uneven and at lower levels, and only once approached $30M (Stampede). This unevenness has keenly affected not only resource availability, but has exacerbated the already substantial uncertainties of competition for sites and staff.

The average capital investment across six years from 2010-2015 is ~$13M/year. There is no surprise that user needs, at least for the great majority of NSF users that do not have access to Blue Waters or DOE systems, are not being met with Track 2 resources! This is the critical issue for users; issues like allocation processes or requirements gathering are second-order.

The operations costs are under-estimated in Figure 1 in the first few years because there were still PACI-era and TeraGrid-era resources being operated under earlier awards. However, especially with prevailing NSF guidelines that annual operations costs must be no more than 20% of capital costs, the trend is clear that operations costs as well as capital costs are diminishing, both in aggregate and for each system. Even with some costs being absorbed by XSEDE, operating costs for recent awards of $1.2M-$3.0M per year are bare-bones, inadequate for the levels of service expected by users, and not sustainable for institutions.

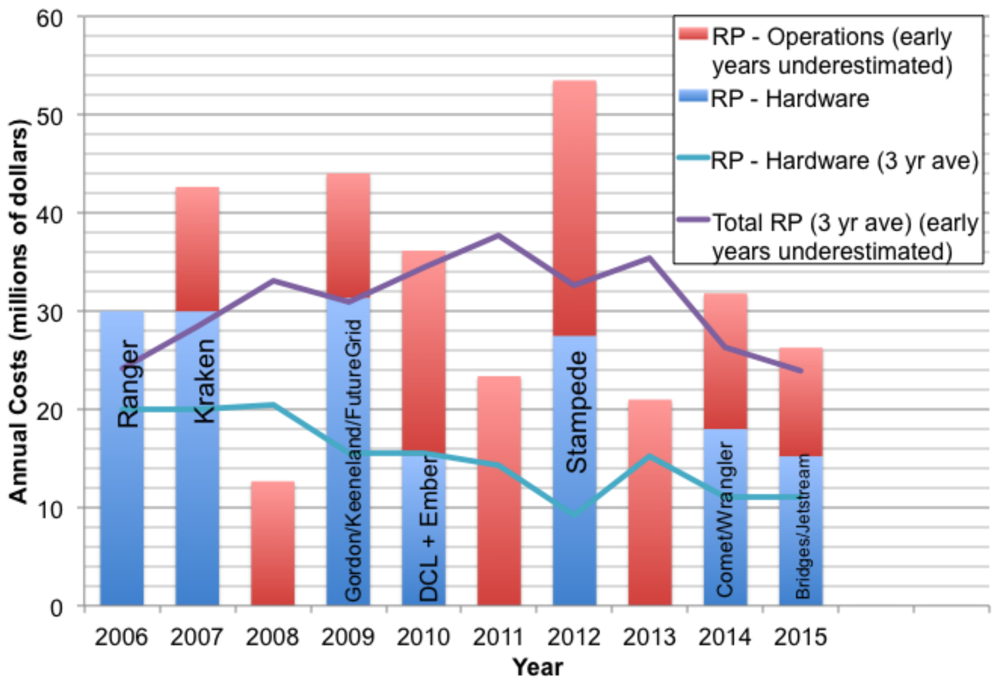

Figure 1. Annual hardware and operating costs for Resource Providers versus time.

3. NSF's investments in HPC resources have become more diffuse.

Hardware investments are split between the leadership-class, Track 1 system (Blue Waters) and several Track 2 systems, with the latter accommodating the majority of the users. Awards per Track 2 system have decreased from $30M in FY2012 for Stampede to $9.6 and $6.6M for Bridges/Jetstream in FY2015. Operational support for the Track 2 systems is further split between small operating budgets for each system and the over-arching XSEDE project.

Recommendations

1. To support an expanding user community, NSF needs to either substantially increase its overall investment or provide more focused, cost-effective investments targeted toward those users.

Providing HPC access for a growing number of users across a diversity of disciplines is an essential component of NSF's support of science and engineering. Many of those users are best served by access via gateways and/or virtual environments. In the absence of a substantially larger investment, NSF should cede the leadership-class systems to DOE to allow adequate funding of Track 2 systems. The latter better match the needs of most users and are more cost effective than commercial clouds.

2. NSF needs to revise its Track 2/XSEDE model to ensure that the systems have enough operational support to be responsive to community needs.

The optimal number of Track 2 systems is a continual topic of debate. However, more small systems will necessarily have higher aggregate operating costs than a few large systems because some operating costs are fixed. In particular, an annual operating cost cap of 20% of capital cost is inadequate for systems with a capital cost less than $30M. We suggest implementing the DOE model, where a small number of labs compete for systems at regular intervals. This retains the competitive nature of procurements, but provides stable career paths for expert staff, whose skills are critical to the successful and widening use of these systems. The final report of the 2011 ACCI Task Force on High Performance Computing also highlights the importance of stable funding for HPC centers, given appropriate oversight processes.