Culture Change and Research Computing at Stanford

Bill Clebsch, Sam Steinhardt, Matthew Ricks, and Phil Reese
IT Services Division
Stanford University, Stanford CA

How Stanford has traditionally managed research and research computing

Stanford University has a long history of being highly rated in national research surveys. A compelling factor in this success has been that each of Stanford's seven schools has autonomy and the associated responsibility of research oversight and funds management. It is unlikely this overall research model will change.

Decentralized support for research computing efforts has largely been managed by the Principal Investigators (PIs) with occasional involvement from the Dean level within the schools. This has led to the School of Engineering having significant research computing resources at the sponsoring faculty and departmental levels but not uniformly across the school. The Schools of Medicine, Humanities and Sciences, and Earth Sciences also have research computing resources available. Meanwhile, an interdisciplinary center, Bio-X, which supports research associated with biology and medicine, has established a shared computing cluster available to those researchers participating in the center.

The modal shift from traditional research efforts to a simulation or modeling basis has researchers either mounting a significant effort to obtain access to existing computing resources or beginning the purchase process for the development of their own resources.

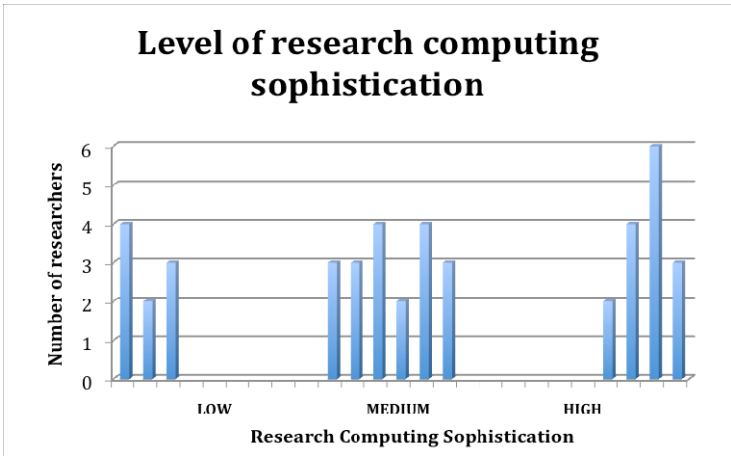
Despite the pervasiveness of decentralized culture, faculty have questioned the practicality and sustainability of each department and/or school mounting its own research computing effort. The faculty with computing systems know their systems are not fully utilized all the time.

To assess the real culture and the opportunity for a change, some 60 interviews were conducted over the past two years, focusing on computing-based researchers, administrators, and faculty. Our conclusion was that almost all of them were considering a modal shift in their research.

Many of the interviewees were neutral regarding where their servers were housed, as long as good network access was provided. Additionally, most researchers would agree to share their resources, in the hope of getting additional resources when needed. Cloud computing is on the radar but not front and center. In fact, the more computationally intensive researchers have found that the cloud model will not work for them.

The long-term recommended approach to address the computing needs of the research community was to build a Research Computing Facility. Then follow up to address centralized equipment and operational opportunities.

One model that evolved from the interviews is presented in the figure, which depicts a continuum of research computing sophistication along the X-axis and the count of faculty at that level of sophistication on the Y-axis. It currently suggests only a slight bias toward the right end of the graph. The interviews show that the more sophisticated group will grow in computing resource needs but their numbers will not change very much. The middle group, whose needs can be met through the use of a shared computing resource, will grow incrementally year over year. The group at the left represents those unsophisticated with research computing who are interested primarily in the science or have little-to-no interest in their computing technology. This is the growth area for campus research computing and likely across the entire research world.



At Stanford, where the whole continuum of researchers exists, none of the groups have all the physical and technical resources necessary to keep their research efforts growing as fast as they desire. Computing cluster size often grows by 100 or 200 percent when a new line of research is undertaken but physical space, power, and cooling needs, distributed across campus, cannot be met.

To cope with this fundamental problem, the Dean of Research and campus IT Services, have proposed a Research Computing Facility building to centralize at least the hosting of the larger computing resources of the schools. This effort is fundamentally a culture change effort.

Private universities rarely mount a new building initiative without a donor providing 25-50 percent or more of the funds for a new building. A data center facility has little aesthetic outward appeal and a high cost per square foot: sponsoring a building like this is of little interest to a donor. This leaves the building development to the university, combining this with the impact of the economic recession will delay the project.

However, the lack of computing cluster space, power conservation efforts, change in the fundamental tools for research, and the overall campus desire to be considered for high profile grants and CyberInfrastructure efforts requiring high end computing have kept the facility effort alive. What is needed is a sustainable business model.

Suggesting a business model for centralized Research Computing

The proposed business model assumes the campus would use debt to fund the construction. By charging indirect costs to sponsored projects and contracts, this initial investment, including interest, could be recouped over time through specific building allocation percentages, that will be based on the proportion of research space to total

space. Furthermore, power costs could be recovered through indirect cost rates, as they are today.

Under this model the ongoing operating and maintenance expenses will be less than when the facilities are distributed across campus, but because they are centralized they stand out as budget numbers that will need funding. The model suggests that the ongoing operating and maintenance expenses would be charged to the Provost's office, which would then allocate space in the facility to Deans of participating schools, based upon their level of financial participation. Each Dean, in turn, would have the discretion either to fund this directly or to charge out to their individual PIs. Some Deans plan to fund this initially and gradually move to charge the PIs.

A fundamental cultural change of this magnitude at Stanford is not undertaken lightly, thus we are launching a pilot project. The existing central administrative data center, built 1978 but upgraded regularly, is being expanded to host more administrative computing. Additional space in the building could be repurposed for a 70-80 rack research-focused data center. The cost for the remodel would be modest compared to that of a new building.

Instead of the traditional 'Service Center' recovery model that the administrative computing side of the data center uses, this space would follow the suggested model for the future Research Computing Facility.

When all goes well with the pilot effort and the available space is put into use quickly, a plan for a full data center will capitalize on the momentum expected from the pilot effort.

Efforts so far have focused on the development of computer hosting facilities. Reflecting back on the continuum model suggests that additional services will be equally important. A centrally provided, efficient and cost effective system support model is the next step.

Make the model sustainable

Researchers planning to use computing resources need to include support costs, above and beyond the hardware costs, when completing grant applications. Clarifying this with researchers submitting grants will aid in the successful change in research modality to a computing-based model.

It is clear that the current accounting rules for the Indirect Cost rate will change as the modal shift in research happens. Getting ahead of that change and leading the way will be better for Stanford than reacting to the changes as they are put into place over time.

While controversial in nature, seeding of a centralized computing cluster resource would create the cultural change faster. Accelerating the process will improve the overall infrastructure efficiency, as well as the research process efficiency. These efficiencies should provide measurable increases in grant support, which could off-set the initial seed money. Lastly, the time from a researcher's initial idea to the time research is completed should be accelerated with this modal change.