Towards a Sustainable Business Model for Campus High Performance Computing

Dan Stanzione
Director
10/3/07

Ira A. Fulton School of Engineering
High Performance Computing

The Givens

- Every campus does some form of HPC.
  - To do research with HPC
  - To do research on HPC (CS)
  - To teach HPC
- We all think we should do big research computing, even if we differ on where and how.
- Everybody pays for it somehow, but there is no consensus how, and many systems end up abandoned (this is bad for everybody).

The Challenges

- HPC has no clear home on campus:
  - It's research infrastructure, so the VPR office might be involved
  - It's computers, so IT should be involved
  - Each now has an excuse to hand it to the other :)
- HPC can be inordinately expensive, and can't be put just anywhere.
- Done even a little wrong, it can be a huge failure.
- "Plain old IT" folks can't provide the right service. "Plain old researchers" can't, either.
- No one knows exactly what HPC is. To some it means Linux clusters, to some it means web servers...
- Every faculty member wants their own cluster.

The Challenges

- Campus IT standards get in the way.
- "Plain old IT" folks can't provide the right service. "Plain old researchers" can't, either.
- No one knows exactly what HPC is. To some it means Linux clusters, to some it means web servers...
- Every faculty member wants their own cluster.
- Systems get old quickly. You'll drop 200 spots a year in the Top500. If you win a grant for hardware money, how will you replace it in 24 months, or will you become irrelevant?
- Can't we get free access from {NSF, NASA, DOE, DOD}?

The Good News

- There are a number of compelling arguments for consolidating HPC funding that have emerged recently:
  - Funding agencies are increasingly concerned about stewardship of their computing investments and data (management plans are often required).
  - Physical requirements of new systems make the "machine in a closet" approach more costly.
  - Increasing campus focus on security makes distributed servers unpopular.
  - HPC is unique among research infrastructure - more on this later.

Guiding Principles in An HPC Business Model

- Free, unlimited services are never successful, and never end up free or unlimited.
- Asking any part of the university for money, and offering nothing tangible in return is a bad plan. Funding from multiple sources is a must. Hardware is the easiest money to find "elsewhere".
- "Self-Sustaining" is a pipe dream.
- The fundamental goal is to increase research productivity (increase $, discovery will follow).
- There is a difference between setting up a cluster, and making HPC-enabled research work. This gap must be filled.

How to Build a Center

- First, acquire allies. You need several types:
  - Administrators
  - Faculty
  - Vendors
  - Some allies are more equal than others; a couple of bulldozers is worth a lot of termites.
- Then, make allies partners:
  - This means everyone has to *put in* something, and everyone has to *get* something (including the administrators).
- Don't start off empty handed...

How To Build A Center

- Third, find some money to get started.
- Figure out how you can provide value (e.g., beyond the machine, what is needed for success?). Get a couple of champions who are research leaders to buy in.
- Plan modest, but expect success. You won't get $10M to start, but you need to be prepared if you do (e.g. if your hardware budget triples due to wild demand, does your model scale the people, or do you go back with hat in hand?). Put in milestones for growth, and make sure all parties will know what success is, and are ready for it. Do the same for failure... under what circumstances should they stop funding you?

How To Build A Center

- Decide what you need from everyone to succeed. Then decide what you can get from them. Try and make these meet in the middle. If you are asking one person for most of the cost, or anyone to contribute forever, you've done it wrong.
- Figure out what you can measure to determine success. Track everything you can (e.g. I know *exactly* what percentage of total university research expenditures come from our users, and what % of proposal dollars involves the center).
- Get faculty buy-in. Figure out what the key faculty need. Carrots work better than sticks, but try and have both. Decide which battles you want to lose, and lose gracefully. The secrets of this don't go on slides...
- Now that you have a plan, be prepared to scrap it at a moment's notice. Be opportunistic and flexible.

...And Then

- The day you get started, you need to start working on your upgrade.
- If you don't know what your next 5 proposals are going to be, you aren't pushing the pipeline hard enough.
- Build a great staff.
  - Your HPC leadership is not the lead admin, it's the person who can write grant proposals.
  - You need someone who has production, HPC expertise. Once you have some of this in-house, you can train more...

The ASU Experience

- Begun with endowment funds
- Housed in the Fulton School of Engineering
- Construction planning began 9/04
- Operations began 9/05
- Year 1 Budget: $600k (Core system 2TF)
- Year 3 Budget: $4,800,000 (Core System 23TF)

ASU Arizona State University
Ira A. Fulton School of Engineering
High Performance Computing

Funding Model

- Core funds: $1.25M
  - University Investment:
    - UTO (IT), Research Office, contribute $250k each annually (40% of total)
  - Three Academic units investing $250k each:
    - Fulton School of Engineering
    - Liberal Arts and Science
    - Life Sciences

The Rest of the Budget

- Core Funds: $1.2M
- Cycle purchase by faculty: $1.2M (80% startup packages, 20% grants).
- Storage sales to faculty: $1.5M (mostly a few large grants, $100k in small projects).
- Direct grant income (awards to center): $900k

Center Offerings

- Obviously, we offer cycles and storage, but also other things...
- Training, sysadmin support, application support, programming, and visualization.
- Grant support, lots of it (ranging from support letters to pushing through the entire proposal; ASU HPC was involved in 40+ proposals this year).
- Help getting national center allocation

Allocations - Cycles

- Three - tiered allocation model:
  - A minimum level of service for free (10k hours)
  - Additional service via proposal to faculty-run allocation committee (100k CPU hours/yr)
  - Services beyond that on *partial* cost recovery basis (there is some creativity here... sometimes by the hour, sometimes by the processor)
- Storage is handled in similar fashion

Where Money Comes From

- Federal grants
- Private Funds
- State Funds
- Research Office
- Central IT
- Academic Units

HPC is important. So is world peace. Why should we get resources?

- HPC is in constant competition with other research/ university priorities
  - IT is already a huge line item in the budget of every university and every unit in it (ASU: $90M, half distributed).
  - It is true HPC is critical to research progress, but somebody will make the same argument about the electron microscope down the hall, or wetlab space, or, sequencers, or...
  - And just because it's important, doesn't mean anyone cares...

What is Central IT After?

- No new problems
  - It is not in the best interest of university IT for each faculty member to run their own cluster and fileserver.
    - Security: IT is still responsible for the network
    - Data Integrity: Who will the Inspector General call when federally funded research data is lost?
    - Faculty will build clusters... IT can either choose to manage or not manage them.

If you don't manage them, they will be broken, underutilized, security problems on which you get lots of calls, and everyone will be unhappy.

If you do manage them, everyone will want different things, it will be impossible to keep track of local state on each of them, and everyone will be unhappy.

What Is the VPR After?

- ROI - Return on Investment
  - Myth: VPRs want you to make great discoveries
  - Reality: VPRs want you to discover you've been externally funded.
- The research office is not a funding agency; a good research office invests in centers that provide a Return-on-Investment. For HPC centers, this means either:
  - Win your own grants
  - Drive research wins (how do you measure impact?)
- Exemplar: TACC

What are the Funders After?

- State Legislatures:
  - Economic Development; states want to create jobs. They will care if you bring in companies or create new companies (employing people in your center does not count as job creation :)
- Academic Units:
  - Some deans will spend some money on this to appease faculty

The Good News: Leverage

- HPC/Research Computing can affect all these areas...
  - Research office is already investing in HPC through start-up packages (on individual faculty clusters... in our case, >$400k/yr)
  - IT is incurring support costs, one way or another
  - Someone is covering facilities costs (find out who)
  - HPC can drive research across a broad range of areas
  - HPC competitiveness focus lately makes ED an easier argument
- It's *not* just another piece of research infrastructure.

A Few Myths

- F&A will pay for this!
  - Research Computing is distinctly NOT in the indirect cost rate... F is physical plant only, and is break even... A is everything else, and is usually a money loser.
- We can set up a cost recovery center! (or faculty are willing to part with money...)
  - This just doesn't work, at least not entirely. Academic centers that tried this are gone. Many commercial services are trying to sell cycles (IBM, SUN, Amazon) without a huge amount of sucess.
- I won an MRI, and the hardware is the hard part!
  - Hardware money is relatively easy, and is actually not the biggest cost. Facilities cost for 20KW racks are huge, and personnel costs forever.

Take Aways

- You need allies to run a center.
- Campus funding is politically complex, but doable.
- Make sure everyone contributes, one way or another, but spread the burden... everyone gets more than they pay for.
- Centralizing HPC on the campus is worth doing.