Improving Research Computing at UCI: the 4-Sigma target

Summary
UCI has a fairly common physical CyberInfrastructure (CI) with which we are trying to attain an uncommon goal: benefiting as much of the campus as possible as economically as possible.

Guidance
While we want to add power to our underlying compute, storage, and networking CI, we are very aware of how that power affects our actual users. A place on the Top500 Supercomputer rankings list means little if that power is opaque to them. Hence, we have a varied group called the Faculty Research Computing and Networking Advisory Group[1] to set priority for IT development campus-wide. Also, our central HPC Cluster is based on the condo model, and thus our Condo Association committee helps to set terms and conditions for the use of this Cluster. This guidance is both critical and appropriate, since we see many instances where dollars committed to CI tend to benefit only a thin swath of users

Infrastructure
Compute: The campus has 2 main compute clusters, one of which is GreenPlanet, a 'strictly walled' condo cluster for researchers associated with the Physical Sciences (PS) only. The other is the HPC cluster which is open to all researchers and is a 'shared condo' cluster; while individual Schools, Departments and PIs can add hardware to the system and have priority on that hardware, when they are not using it, the hardware is open to all users on a sliding scale, with other hardware owners having priority before unaffiliated users. This assures that all users get more cycles than they paid for, while also allowing users who cannot afford the buy-in costs also get a fair amount of compute cycles due to the spiky nature of owner usage.

Storage: Most Schools and departments provide small (~TB-sized) file storage for their users, mostly CIFS/SMB and Central Administration provides small amounts of highly hardened storage for their use on NetApp appliances. The 2 compute clusters provide storage in distinctly different ways. Both HPC and GreenPlanet provide NFS mount points for user-provided storage hardware, but HPC also provides 2 large Distributed File Systems (DFSs), for high bandwidth research computing. These DFSs are key to providing fairly cheap, robust very high performance storage.

Networking: The UCI networking is fairly standard for a university of its size. It is described further below.

Support: UCI has a small but experienced Research Computing Support staff, supplemented by additional School and Departmental support staffs. With the exception of PS) and Information and Computer Sciences (ICS), the support staff handle exclusively desktop and administrative issues. PS and ICS have their own Data Center staff and are trained in technical computing. The central Office of Information Technology (OIT) Research Computing Support (RCS) staff include 3.5 FTEs, of which 2 are PhDs (in Molecular Biology and Signal Processing) and RCS handles by far the largest and most diverse computing facility, the HPC cluster.

The 4-Sigma target
Many CI plans and funding target the upper 1% of research computing users, whose requirements put the current network under strain with extreme demands on network bandwidth, latency, storage, and compute requirements. While addressing the needs of the 1% sometimes lead to improvements for the 99%, often it does not. Our immediate CI plans are to improve computational resources for the middle ~95% of researchers – the ± 2 Σs of the research computing population – the 4-Sigma in the header. By making it easier for most of the population to do their work with an overall infrastructure change, not only will they
become more productive, but there will be more time to dedicate support to the top 1% who pose the most challenges (but who are also often the best positioned to help themselves).

**Seed → Sustain**

Generally, there are two phases to any major upgrade in infrastructure. The first is Seed funding; the second is Sustain funding. Seed funding usually derives from a major administration-funded initiative or from a grant. Sustain funding is usually more problematic, since it is hard to convince any group to contribute and to convince them that what they are funding has perceptible return to them. However, we have good reason to believe that using the Condo model can convince a majority of stakeholders to fund this approach. Our compute clusters have been successfully funded via the Condo model, in which users contribute self-bought resources to a common pool and the resources are distributed to all users based on contribution and immediate need or willingness to wait for a lower priority resource. This creates a pool of resources that can be allocated in a flexible way and tends to buffer spikes of usage and create more apparent resources for everyone.

**Storage**

From repeated faculty surveys, the most requested resource is storage for faculty research. The details vary as to what kind of storage is most desirable (back-ups, ease-of-access, reliability, high performance, web-available, for active data, or for archiving), but it is always storage that leads the list. The Condo model for storage has the University seed a small storage cluster and provide the Data Center space, system administration, and networking support. Interested parties would pay to expand the storage, and according to their requirements and their contributions, would be added via automatic quotas. Typically, smaller groups buy excess storage to allow for growth, but this is wasted money since the cost for disk-based storage halves every 14 months, and we can use a scalable DFS to increase storage only as needed. Therefore, instead of carrying a large overhead of unused storage, we keep the storage pool at just under critical levels, increasing it in ~50TB chunks. This increase is fairly cheap since we use high-quality, but Commercial Off-The-Shelf (COTS) technology.

Such a large storage pool can be used for a number of purposes, depending on configuration and networking, but networking is the critical point for this application. With a recent campus network upgrade to 10 Gigabits/s (Gbs), we can use this campus storage pool for both general purpose storage as well as fairly high performance storage that connects directly to the campus compute clusters and to other high-bandwidth sources and sinks, such as genomic sequencers, imaging machines, remote sensing repositories, and other Internet-based data archives that provide both static and streaming data.

**Networking**

A top-tier research and teaching university requires first-rate networking to support data-intensive research and, increasingly, multimedia-based teaching. Multimedia-based teaching can efficiently expand the scope of the UCI’s pedagogy.

The UCI campus network (UCINet) consists of ~1700 network routers and switches with approximately 40,000 active network connections in 175 buildings on campus. There is also a campus-wide WIFI network providing wireless connections to the campus community. UCINet has a 10 Gbs fully redundant backbone linking all buildings together. Thanks to our recent NSF-funded upgrade, 11 campus research buildings have a 10 Gbs uplink connection to the campus backbone and the rest of the buildings have 1 Gbs connections. However, only two buildings (Calit2 and Bren Hall) have 100% 1 Gbs connections to end users; the others have 100 Megabit/s to the majority of end users.

The campus network is currently connected to the Internet via our Internet Service Provider, CENIC, with a total capacity of 20+ Gbs. Among them, one 10 Gbs link is to CENIC’s High Performance
Research Network (HPR). CENIC HPR has direct 100 Gbps connections to National Research & Education Networks, such as Internet2 and Energy Science Network (ESNet).

**Bandwidth Upgrade within Campus**
To allow scientists with high data rate requirements to be productive, and to attract and retain the best faculty, we have (via the previous NSF CIE):
- Increased bandwidth to selected research buildings (Rowland Hall, BioSci 1 & 2) from 1 Gbps to 10 Gbps.
- Increased bandwidth to 1 Gbs to the Desktop for researchers who perform data intensive applications by installing a 1 Gbs switch on each floor of the selected buildings.
- Enhanced bandwidth capacity of campus wireless core infrastructure by adding 10Gbps capabilities.

**Bandwidth Upgrade for UCINet External Network Connections**
In order to share BigData at reasonable speeds with remote sites and to take advantage of grid-distributed resources such as the Large Hadron Collider Tier1 and Tier2 data nodes, we have:
- Added a 100 Gbps connection between UCINet and CENIC’s HPR Network.
- Upgraded the campus border capacity to accommodate the upgraded external connection.

**Networking Performance Improvement at the OIT Data Center (OITDC)**
To improve the network performance for our large clusters, which contain most of the campus research storage and servers and resides in the campus main data center, we have:
- Upgraded connection bandwidth to 10 – 40 Gbs.
- Replaced low performance switches.
- Increased fiber capacity from ADC to the campus CORE sites.

**Support for New Technology**
- **IPv6:** An IPv6 /32 address (prefix: 2607:f538::/32) is allocated to UCI by ARIN. Deploying IPv6 on campus will require the upgrade of the eight-year-old backbone routers and many building routers. The backbone router refresh is included in the equipment refresh plan described in section 3.2.4.

  Once that is done, implementation of IPv6 can take place. We have not received requests for IPv6 from our user community, but when such needs arise and before the general deployment, we will do our best to help these individual cases to achieve their goal.

- **Software Defined Networking (SDN):** Openflow based SDN will be implemented in accordance with researcher requests. We will make sure all new purchases of critical network equipment is SDN capable. We will also actively exam the campus network and data center to see where SDN implementation would benefit the campus community the most.

- **InCommon Federation:** UCI is already a member of the InCommon Federation, which uses a standard protocol for establishing trust relationships. We subscribe to the InCommon Digital Certificate service and intend to be certified in the InCommon Assurance Program at the Bronze and Silver levels.

**Human Resources**
One of the problems with the increasing rate of change in the digital workplace is that human ability often lags the state of the art. To address this problem, UCI’s CI plan involves improving the human resources available to assist with these problems, especially with the research aspects. We will be increasing the available HR to support:
- **Scientific Programming:** With the increased amount of digital data, there is widespread desire to mine and analyze it efficiently. We are now well past the point where these datasets can be manipulated with Office applications; the tools of Big Data are required to exploit them. To this end, UCI needs people who can know these approaches and toolkits, can make them available to
researchers, and help them get their workflows working. This is specialized knowledge and we are competing against Google, Amazon, and the like to hire and keep such people.

• Storage and Archiving: This is generally the purview of the Library system, but especially due to many granting agencies requiring specific types of archival storage and integration with specific retrieval software, there is a need for either more training of the Library personnel or hiring more people who have that expertise. This overlaps with other forms and uses of storage mentioned previously.

Hardware
While this is an ongoing requirement, a number of people are required by their jobs to keep up with hardware advances. So, while it would be nice to have hardware-specific experts, this is one area where we are not as stressed as in others.

Software
Since CI is at base a software-mediated infrastructure, careful policy and review at this level is among the most critical decisions that can be made. Software has huge possibilities for improvement: in robustness, capability, and especially scalability. Over the past few years, the number of mature software packages available as Open Source Software (OSS) has increased by orders of magnitude. While there are still a few proprietary software packages that have no corresponding OSS equivalent, the OSS world is rapidly reducing that disparity. This is especially true for anything that acts as a server; there is an excellent chance that there is OSS that is as good, or better than, the proprietary version.

While it may not be possible to replace existing proprietary packages with their OSS equivalents due to the long tail of entanglement that commercial packages encourage, we will use Open Source packages going forward where there is not a compelling (and documented) reason to use the proprietary one. This is an especially notable step since University of California has 10 campuses and good channels to spread information across them. We should be, and will be, using these channels to provide best-practices recommendations and especially configuration of mature OSS packages for particular uses.

Scalability
We are always looking for ways to make CI more scalable. We have done it well with our compute cluster architecture, and we will re-use the same techniques to approach a campus storage system. While hardware can only scale so far, it is also continually falling in price, so we are able to buy more capacity for less money over time. Hardware also has a fairly short lifetime: typically 3-5 years between refresh cycles.

Configuration
Software is different. Unlike hardware, it tends to persist for a decade and longer since it often takes a huge amount of effort to make it work well and once it works well, it tends to persist until major infrastructure changes can no longer support it. Because of this long tail, decisions regarding software policy are critical. Some software (both proprietary and OSS) can be troublesome to install, but overwhelmingly the real time cost is in the configuration. Once a configuration has been set, it is much easier to scale out the deployment for a particular package, especially if an application has a single configuration file and allows in-line comments as do many OSS packages. If it takes a day to configure an application to work well, but that investment of knowledge can scale over 100 OSS deployments, then the overall time per deployment is trivial. As more services are based on server offerings, either as cloud services or via an institutional server, the time per configuration reduces dramatically. This is what we are seeing.
Cluster-based Instruction
Most lab-based computer instruction still takes place in a classroom, mostly because of the need to have specific client machines licensed for particular software. However, as instructors find OSS that fulfills a particular need, they are requesting server-based support for running the software for classes. Currently UCI is using externalized compute nodes from our HPC cluster to provide this on an ad-hoc basis, but going forward we will have to formalize this approach to provide instructional computing. This allows classes themselves to be distributed if desired and OSS remote display software such as VNC, NX, and x2go allows students to connect to Graphical User Interfaces generated on the server, even if they are separated from the server by 10 or more network hops.

This server-based approach reduces the requirement of supporting computer labs since almost all students have their own laptops and can use the server software from anywhere. This seems to be the future of lab-based computer instruction, although the one area that still needs work is software that requires hardware-accelerated graphics.

Licensing
This is a continuing problem since most proprietary software companies have different licensing requirements, methods of software distribution and updating, mechanisms of licensing, discounting by user numbers, licensing periods, optional add-ons, prorating of licenses, and support mechanisms. For all of these reasons, it would simplify things significantly if we could use OSS. There is an ~2 FTE/campus cost to just administer the licensing and manage the various license servers and the software configuration itself. At least 1 of these FTEs could be switched to direct user support.

References

1. Faculty Research Computing and Networking Advisory Group:
   Description: <http://sites.uci.edu/rcna/>
   Membership: <http://sites.uci.edu/rcna/committee-members/>