

LSU Cyber-Infrastructure

Summer 2013

Compute Cycles

Present – For more than a decade, LSU faculty and students have had access to significant high-performance computing (HPC) resources at no charge to users. Currently, the two systems that supply the majority of computing cycles are *SuperMike-II* (a 212 TeraFlops cluster that came into production at LSU in February, 2013) and *Queen Bee* (a 51 TeraFlops cluster that was acquired in 2007 as part of LONI, the Louisiana Optical Network Initiative). Details regarding the architecture of these two HPC systems can be found online, but it is worth noting here that 52 of the 440 compute nodes on SuperMike-II contain a pair of NVIDIA ‘Kepler’ GPUs. With funding from a 2012 NSF/CNS award, LSU also has recently installed *Shelob*, an HPC cluster in which all 24 nodes contain 3 NVIDIA ‘Kepler’ GPUs. The GPU nodes on *SuperMike-II* and *Shelob* offer LSU faculty and students an avenue for accelerating targeted computational tasks locally as well as for developing codes that can be effectively ported to even larger HPC platforms at, for example, NSF and DOE supercomputing centers.

Near Future – LSU anticipates bringing two new HPC systems online during the 2014 calendar year: (1) A \$4M NSF/MRI award, along with matching funds provided by LSU’s CCT and Office of Research & Economic Development, will permit us to build *SuperMIC* – a 380-node, Intel Xeon Phi-based HPC system with a peak performance of ~ 1 PetaFlop. (2) Funding provided by the Louisiana Board of Regents to LONI should allow us to replace Queen Bee with a system whose peak performance rating is 2-3 PetaFlops.

Digital Networks

Present – For more than a decade, LSU has provided its faculty and students with excellent digital network connectivity in support of research activities campus-wide. A high-capacity fiber optics network interconnects all LSU buildings and, bolstered in 2010 by a \$2M NSF/ARI-R² award to “revitalize LSU’s data network infrastructure,” digital routing equipment supports 10 Gbps or greater connectivity between major research buildings with 1 Gbps feeds into multiple laboratories in each building. LSU’s external connectivity is facilitated through *LONI*, a Louisiana Board of Regents initiative funded at a level of \$5M per year, which – in addition to supporting the HPC resources mentioned above – interconnects all of Louisiana’s institutions of higher education and ties those institutions into Internet2 and the National Lambda Rail via multiple 10 Gbps links. The core hardware routers and technical staff for LONI, as well as for the LSU campus, are housed in LSU’s Frey Computing Services building along with SuperMike-II. With \$0.5M in funding from a 2012 NSF/CC-NIE award, code-named CADIS, a 100 Gbps digital network pipe is being constructed between the Frey and LSU’s new Digital Media Center (DMC) building in order to support the wide range of compute- and data-intensive projects that are pursued by CCT researchers.

Near Future – In concert with the new customer service model being developed by Internet2, LONI is gearing up to support one, and possibly two, 100 Gbps connections into Internet2. In addition, a recently announced, \$1M NSF/CC-NIE award will allow LSU to develop a software-defined network (SDN) that interfaces seamlessly with LONI’s “science DMZ” to support multiple big data projects across the campus.

Data Storage & Management

Strategy – LSU is implementing a strategy designed to ensure that the university’s cyber-infrastructure includes sufficient high-capacity and long-term storage to support the “big data” research and educational activities of LSU faculty and students, now and into the future. In so doing, LSU will

satisfactorily respond to the data management mandates that are being issued by federal funding agencies. The strategy has been developed around the following key considerations:

- The amount of storage should not only meet the present needs of LSU's research and educational programs but should be readily expandable to cover future needs.
- The storage should be multi-tiered – 3 distinct tiers are identified, below – to take advantage of cost differentials associated with different hardware solutions while accommodating a range of access needs.
- The solution must accommodate secure (e.g., HIPPA compliant) as well as unsecure and publically accessible (i.e., web servable) data.
- Adequate levels of disaster recovery services must be provided.
- The solution must include a curation layer, that is, a software interface that adds data/file identification (metadata tags) to facilitate search and retrieval operations.
- Technically qualified IT and library staff should work as a team to provide essential levels of maintenance, access, support, and archival curation.

Present – LSU recently has purchased and installed a base level of data storage capacity to support a broad range of activities among its researchers.

- Tier 1 storage (high-performance disks in a lustre/linux-based environment with a high-bandwidth, parallel interface to LSU's HPC system, SuperMike-II): Total available is 1 PetaByte with a 4-year maintenance contract; cost is ~ \$125/TByte/year.
- Tier 2 storage (disks integrated with LSU's robust administrative storage solution; files are straightforwardly CIFS-mountable on PCs and Macs): Total available is 100 TeraBytes with a 3-year maintenance contract; cost is ~ \$800/TByte/year.

In the fall of 2013, all tenured or tenure-track faculty at LSU (approximately 1000) will receive a tier-2 storage allocation of 100 GBytes/faculty. Beginning at that same time, faculty with large “compute cycles” allocations on SuperMike-II will be granted access to the Tier-1 storage; there are approximately 50 such groups so, on average, they can expect to receive a tier-1 storage allocation of 20 TBytes/group.

LSU is committed to maintaining – in perpetuity, for its faculty who are active in research – at least the base-level storage capacity that is identified here in connection with Tier-1, Tier-2, and (see below) Tier-3. The cost associated with this commitment is estimated to be \$0.5M per year. In addition, all three Tiers are readily expandable, so researchers may use extramural funds to annually secure larger storage allocations within any one of the three Tiers.

Near Future – By summer of 2014, LSU plans to expand its base level of support for research-related storage to include ...

- Tier 3 storage (archival storage that will ensure permanency of files containing key research results; storage capacity is emphasized over speed of access; could be tape robots or non-spinning disks): The goal is to allocate to each research-active faculty member 2-5 times the storage of Tier 2, but at a fraction of the cost. This might be achieved by contracting with commercial cloud services; estimated rate is \$720/TByte/year.

LSU expects to add a data curation layer to its Tier-2 – and, ideally, also Tier-3 – storage offerings and, in so doing, provide the basic CI required to establish an *Institutional Repository*. Software tools and interfaces being examined include DataVerse (<http://thedata.org>), iRODS (<https://www.irods.org>), Alfresco (<http://www.alfresco.com/tour>), and Purdue University Research Repository (PURR; <https://purrr.purdue.edu/>). LSU also is investigating providing file management tools that permit straightforward and routine file migration between all three storage tiers.