

REPORT DOCUMENTATION PAGE

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing existing information, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 02/09/2010		2. REPORT TYPE Final		3. DATES COVERED (From - To) 05/15/2008-11/14/2009	
4. TITLE AND SUBTITLE (DURIP-08)A Second Generation Flexible Computing Environment for Computational Modeling of Brain Function and Neuroimaging Data Analysis				5a. CONTRACT NUMBER FA9550-08-1-0245	
				5b. GRANT NUMBER FA9550-08-1-0245	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Cohen, Jonathan D.				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Trustees of Princeton University Office of Research and Project Adm 1 Nassau Hall Princeton, NJ 08544-0001				8. PERFORMING ORGANIZATION REPORT NUMBER 349-6014	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) USAF. AFRL AF Office of Scientific Research 875 N Randolph St. Room 3112 Arlington, VA 22203 RSL				10. SPONSOR/MONITOR'S ACRONYM(S) USAF.AFRL	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The instrumentation purchased with funds from this Durip grant was part of a total upgrade of the computing facility of the newly formed Princeton Neuroscience Institute (PNI, formerly the Center for the Study of Brain, Mind and Behavior). This proposal to the Air Force Office of Sponsored Research's (AFOSR) Durip program was intended to complement a parallel proposal for a file server upgrade that was submitted to the National Institute of Health (NIH). Both proposals were funded, and AFOSR permitted the PNI to submit a revised budget that redistributed the Durip funds, reducing the amount spent on the file server (as that was primarily covered in the NIH grant) and increasing the amount spent on the computational cluster. The result was that the PNI was able to procure a new computational cluster and large mirrored file server, resulting in a total facility upgrade that is now sufficient to support the ever growing needs of the PNI for the next three to five years.					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 5	19a. NAME OF RESPONSIBLE PERSON Jonathan Cohen
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code) 609-258-2696

20101022181

Instrumentation purchased

File Server and Backup System

The file server is a mirrored system with two 100TB units from Exanet, Inc. – one located in Green Hall where the PNI neuroimaging research facility is housed, and the other off-site in the campus computing facility at 87 Prospect Avenue. At the time the proposals were written the expected vendor was BlueArc Inc., vendor of the 9TB file server that was already in use.

However, once the entire scope of the project was funded an extensive exploration of a variety of NAS vendors was undertaken to find the best overall architecture for the integration of both the file server and computational cluster. Due to significant competition between vendors and the normal decrease in disk prices over time, we were able to procure considerably more disk space within the limits of the file server budget – 200TB of usable space divided between two mirrored server systems, versus the originally proposed 87TB of just one.

This mirrored architecture is ideal, providing sufficient disk space for a complete copy of the data from the primary server to an off-site location. The mirror to the disaster recovery (DR) system occurs nightly through a high-speed (10Gb) network connection. The primary file server makes use of an extremely high performance, enterprise level disk array from Data Direct Networks (DDN), while the mirror site uses a slower, less expensive, array of Xyratex disk drives.

Also purchased was a UPS from Eaton Corp. for the primary (DDN) disk array that is sufficiently sized to handle the current primary disk array as well as retaining capacity for future expansion. The University's computer room, where the DR system is housed, is entirely on a UPS so a separate unit was not needed.

The final component of the file server system is a 300-tape library, manufactured by Qualstar Corp., with two LTO tape drives, also residing in the computer room at 87 Prospect Avenue. The NetVault backup software (already licensed to the PNI) is used to manage backup tasks and the tape library, which was partitioned to provide both a third level of data protection as well as archival space for data and projects that no longer need to be stored online.

Computational Cluster

The new cluster, from SuperMicro Computer, Inc., fits in a single 36U rack and consists of 13 units each containing four servers, for a total of 52 nodes. This architecture allows high density and power savings since all four nodes in a unit share a single power supply. Each node contains two quad-core Intel Xeon X5570 processors, operating at 2.7GHz, or eight cores per node, for a total of 416 CPUs. This is a substantial increase over the former cluster that was comprised of 64 Apple G5 dual-core processors, both in terms of the number of processors and the dramatically better performance of each. Each node contains 24GB of RAM for a total of 1.25TB RAM. The nodes communicate with each other via DDR (20Gbps) Infiniband (Mellanox) in a flat, non-blocking network. The nodes interconnect via a 72-port DDR Infiniband switch, from QLogic, with 52 ports funneling into eight 10Gbps uplinks to the file server, with 12 open IB

ports available for expansion. The cluster nodes run the 64-bit Linux operating system CentOS 5.3.

Migration

Considerable effort was devoted to the task of seamlessly migrating users from the old computing facility to the new one. Data was migrated first, as the new file server was operational while the search for the computational cluster was underway. Data was migrated one lab (filesystem) at a time, following the same naming conventions and mount points. During transition data was available on the old server in a read-only state, and once a lab's data was migrated, the old mount was disabled, replaced by read-write access to the data on the new file server. Migration of smaller file systems were completed within a day or less, and larger labs were migrated over weekends. The entire process took several weeks, with minimal disruption to users and their research.

It was not possible to power both the old and new computational clusters at the same time due to power and cooling constraints of the computing facility. A subset of nodes on the new cluster was brought up in a temporary location (it only had sufficient power for three of the eight units, or 64 processors) for the purpose of system configuration and porting of software packages, user applications and scripts. Software installed on the new system is functionally the same as that on the old computational cluster. 64-bit Linux versions were downloaded and installed, and scripts were ported and tested. The change from OS X to Linux has resolved some issues with graphics, which has actually improved the graphical interface capabilities of the cluster. Users had approximately one month after initial system configuration and integration with the file server to test their programs and scripts on the new cluster before the old cluster was taken out of service. As of January, 2010, the new computational cluster was fully operational.

The full computing facility upgrade was completed in January, 2010 with virtually no downtime for the users and minimal impact on the PNI and ongoing research.

Impact on user community

The impact on the user community was felt immediately, as the severe crunch for disk space has eased, and there is now room for continued expansion. Laboratories have been able to merge data stored on several file servers onto one resource pool, simplifying their daily data management activities and data sharing needs. Laboratories are now able to run analyses that generate sizeable amounts of data that were previously precluded or more limited in scope due to insufficient storage.

The new file server is also substantially faster than the prior generation's BlueArc that it replaced, facilitating research by allowing jobs and applications of every type to run faster. Also, importantly, the new file server provides total data redundancy, to within one day, for a drastically improved disaster recovery scenario, both in terms of data integrity and speed of recovery. The redundant systems will also reduce downtime, as it is a simple process to switch user access from one file server to the other if necessary.

The user community has experienced a significant increase in compute power with the new cluster. Taking a classic neuroimaging analysis task – correlating a stimulus with the fluctuations in brain activity as measured by fMRI – a single core on the new cluster is 3.7 times faster than a single core on the old cluster on the AFNI benchmark. Although the new cluster has 3.3 times more processor cores (416 versus 126), we are experiencing a 12-fold reduction in compute time when performing identical tasks due to a significant increase in processor speed. Utilizing a network that is two orders of magnitude faster (10Gbit versus 100Mbit), and a system with a 3-fold increase in memory per core, from 1G to 3G, time spend loading and saving data has dropped by at least a factor of 10 in practice.

Administration, operation and maintenance plans

Extensive renovation of a space in Green Hall to house the primary system and UPS included entirely new electrical circuitry and the installation of a 3-ton air conditioning unit. The DR system and tape library are sited in the University's main server room at 87 Prospect Avenue, which is entirely on a UPS system. A high-speed, 10Gb network infrastructure connects both systems to the campus network for user access, and to each other for nightly mirroring. The file server is available to users 24/7. Hardware redundancy, RAID, and mirroring all contribute to extremely stable environment for users, and limited downtime due to either system problems (such as failed disks or fans) or upgrades.

The files are administered by a staff of three system administrators with extensive experience on the previous file server appliance, as well as many years of employment with the University. The primary system administrator has been with the University for over 18 years, and has been with the PNI since its inception in 2005 and, before that, with the Center for the Study of Brain, Mind and Behavior since 1998. PNI administrative support staff manage the day-to-day operation of the Institute and, with regard to administration of the computing facility, manage purchasing and supplies and other operational functions.

Hardware and software service contracts were purchased for all of the new computing facility equipment. In some instances, three or five year contracts were purchased outright, while the main service contracts for the file servers and disks are renewed annually. Routine maintenance will be conducted as recommended by the equipment vendors with no or limited downtime, depending on the type of service required. Downtime is usually scheduled for no more than half a day, and users are given sufficient notice and an opportunity to request a different time should there be a deadline or other conflict.

Research accomplishments

The file server has provided the space needed for several large datasets in its first few months of operation. Uri Hasson's lab (Psychology) has gathered datasets of subjects viewing several movie clips, as well as resting-state data in which subjects lie in the scanner for up to an hour at a time. During this time the scanner is gathering a full 3-dimensional high resolution image every two seconds. Peter Ramadge's lab (Electrical Engineering) in collaboration with Ben Singer (PNI) has begun testing functional connectivity software on existing data sets, storing large connectivity matrices (100,000 voxels by 1500 timepoints for 10 subjects) for each of

several candidate algorithms being considered. Multivariate Pattern Analysis (MVPA) Toolbox users (Norman, Botvinick, Cohen labs in Psychology) have sped up spherical searchlight algorithms, which requires analyzing different regions of interest within large datasets – each region, which will eventually reside in memory to be delivered with our new cluster, can be now saved to disk, albeit temporarily, to increase performance on memory-starved older machines without fear of exceeding disk quotas as has historically plagued researchers in those labs.

These are just a few specific examples of research accomplishments that were not possible without the additional disk space that the new file server makes available. Also significant to note is drastically improved access speeds to data due to the new technology and underlying 10Gb network infrastructure.

The new cluster has been operational for just one month, and is already playing a central role in PNI research. The Norman lab is running multivariate pattern analysis (MVPA) algorithms on the cluster that classify fMRI brain activity in less than half the time needed on the old system. The Cohen lab is expanding the parameter space of simulations that add top-down modulation to MVPA classification algorithms. The Ramadge, Hasson and Turk-Browne (Psychology) labs have been able to correlate fMRI activity both within and between a greater number of subjects than previously feasible, forming a more complete series of functional connectivity maps than was possible before. There are just a few examples of work on projects that were in place when the cluster arrived.

Users are just at the beginning of exploiting the new resources at their disposal. In addition to being able to do traditional analysis much more quickly, we expect researches to branch into new kinds of analyses not previously available due to the constraints of the old facility. Speed increases of this magnitude should open new frontiers in research such as real time brain pattern classification and interactive data visualization.