

Gerstein (Bass Central/Main campus)

The Gerstein laboratory is found in two connected buildings. The laboratory consists of 6 rooms and comprises a total of ~1,900 sq. ft. In addition, three conference rooms that have projectors provide venues for interaction. There are 40 gigabit-ready desks, equipped with one or two 23" and 30" LCD screens. The space is properly air conditioned for supporting a large number of computers.

Office:

Mark Gerstein's office space is 178 sq. ft.

Gerstein Lab Computer Infrastructure

Laboratory Network and Storage. The lab's computing infrastructure is partitioned into a private and a public network. The entire infrastructure is fully gigabit capable and is connected to the Yale backbone via gigabit optic fibre; the network architecture was designed with computing efficiency and network security in mind. The private network consists of individual laptops, desktops and workstations, as well as communal computational servers, dumb terminals, a central fileserver, a consolidated NAS, and printers. There are also servers that provide essential network services such as NIS, NFS, SMB, DHCP, monitoring and backups. The public network consists of numerous production webservers that are either real or virtual machines. The laboratory maintains its own public subnets of 128 public IP addresses and manages many of its own domains (e.g. gersteinlab.org, molmovdb.org, pseudogenes.org, and partslst.org). The lab has a full-time administrator maintaining the network.

The private and public networks obtain gigabit connectivity through four HP Procurve 5300xl switches that are mutually connected via fibre. The private network is behind a Cisco PIX 525, which is concurrently used as an IPSec VPN gateway into the private network. Within the private network are two NetApp storage appliances with 43Tb of raw space, which is configured with 27.5Tb of working space, thirty custom made 4Tb network disks with a total 120Tb capacity, a Dell NAS with a total of 30TB capacity; the NetApp appliances and Dell NAS are used for live user file space, backups of user files and backups of public production webservers. A seven-day incremental backup and a twelve-month incremental backup are currently being implemented in the lab.

Wireless access is available all throughout the lab. Wireless access connects computers directly to the public network.

Available Computers. There are about forty-seven working laptops in the lab, in which eighteen are recent Macbook Pro models.

In total, the lab has 315u of rack space spread over eight racks. Residing in these racks are a dual CPU twelve core Opteron server with 256GB of memory, a dual CPU six core Opteron server with 128GB of memory, a dual CPU four core Opteron server with 64GB of memory, three Intel blade enclosures with 10 dual CPU Intel blades each, fourteen dual cpu 64 bit Xeons servers and six dual cpu 64 bit Opteron servers; these rack servers are in addition to the NetApp storage appliances and the Dell NAS mentioned above. The rack servers have various uses. The dual CPU Opteron servers are for hosting virtual machines, which function as web hosts. In the private network, five rack servers are for essential network services, four are storage head nodes for the Dell SAN and a few are network support or experimental machines. The rest of the rack servers are in the public network acting as webservers. The private network has seven business class color laserjet printers.

Software. A number of open source software, programs created in-house, and proprietary software is used by the lab researchers for their needs. The lab maintains a set of wiki servers for the documentation of internal information and the public dissemination of information. The lab also manages mailman servers for its mailing

lists. The compute nodes are mainly used to develop and run Java and Perl code and to perform Matlab and Gromacs calculations. The public webservers are used to deploy Java, Perl, PHP and Python applications.

Individual tasks are coordinated by a web group calendar. Web applications and servers are continually being monitored by a Nagios monitoring system.

Yale Life Sciences Supercomputer. The Gerstein laboratory has priority access to two of the Yale supercomputers, namely Louise and BulldogI, and regular access to six other Yale supercomputers. There are two full-time administrators maintaining the supercomputer.

Louise is a cluster with 112 Dell PowerEdge R610 with (2) quad core E5620 nodes, each with 2.4 Ghz cpu cores and 48 GB RAM. They are interconnected with a Force10 network switch. There is therefore a total of 112×8 cores = 896 cores. Louise has 300 TB (raw) of BlueArc parallel file storage.

BulldogI is a cluster consisting of a head node and 170 Dell PowerEdge 1955 nodes, each containing 2 dual core 3.0 Ghz Xeon 64 bit EM64T Intel cpus, for a total of 680 cores. Each node has 16 GB RAM. The network is Gigabit ethernet. BulldogI runs a high performance Lustre filesystem. It is managed via PBS. Three 20Tb Dell Power Vault with storage arrays are attached to BulldogI and are dedicated for Gerstein laboratory use.

The laboratory also has priority access to a SGI F1240 system. This system has 12 Xeon E5345 Quad-Core 2.33GHz CPUs (for a total of 48 processor cores), with 2 x 4M L2 cache per CPU, a 1333MHz front side bus, 96GB of memory, and 6 Raptor 150GB, 10K rpm SATA drives. It runs SUSE Linux Enterprise Server 10 as a system single image. That is, all 48 cores are managed by a single process scheduler, and the 96 GB memory is, in principle, addressable by a single process. In practice, system caches and buffers reduce the maximum amount of memory available to any given process to about 70 GB. In many ways then, the system can be thought of as an SMP, but in terms of hardware architecture it is closer to an infiniband-connected cluster.

Core Lab. The Gerstein Lab is adjacent to the Yale Center for Structural Biology (CSB) Core laboratory. The Core laboratory resources are available to members of the Gerstein lab. The Core laboratory supports the work of all the people associated with the CSB, in total about 200 users and >200 computers. These computers include a number of high-performance graphics workstations for visualizing macromolecular structures and complex data sets. The CSB Core staff of 2 FTE provides support to the associated CSB laboratories as well as the Core computers.

Oracle Server. Yale University has an institutional site license for the Oracle database management system.

As a result, many major administrative computing systems at Yale are being developed using Oracle, and Yale's ITS staff has extensive Oracle experience. Yale ITS maintains and operates several Oracle database systems at the School of Medicine, and provides access to these machines to many different projects. There are several advantages to using institutional servers. The ITS staff backs up each database on a regular schedule, typically with full backups weekly and partial backups several times a day. The ITS staff maintains the hardware of the database machine, the system software, and the Oracle software. They perform periodic upgrades when new versions of the software become available. They also handle any systems problems that occur, and are available to help troubleshoot any application problems that arise.