The UNC Department of Biostatistics was recently ranked the 4th highest Biostatistics Department in the nation. It has 50 Faculty members, 68 permanent staff members, 130 temporary employees, and 184 students (165 graduate students and 41 undergraduate students). The Department of Biostatistics occupies a total of 32,099 square feet in three locations and has a state of the art conference room located in McGavran-Greenberg Hall, which seats 20 people. It has a drop-down projector, wireless capabilities and conference calling facilities. Dr. Hudgens’s office is in McGavran-Greenberg Building. The Department of Biostatistics contains numerous biostatistics cores for other UNC centers and institutes, such as the UNC Center for AIDS Research (CFAR) Biostatistics Core, the Lineberger Comprehensive Cancer Center Biostatistics Core, the North Carolina Translational and Clinical Sciences (NCTRaCS) Institute Biostatistics Core, the Center for Environmental Health and Susceptibility (CEHS) Biostatistics Core, and Epidemiological Methods Facility Core. The Core also has close ties to statistical science professionals who are located in other departments and schools, including: the Departments of Statistics and Operations Research; Psychiatry; Epidemiology; and Genetics; and the Schools of Dentistry, Nursing, Medicine, and Pharmacy. The UNC Chapel Hill campus is a rich source of biostatistical resources, such as the extensive holdings and online services of two specialized libraries: the Health Sciences Library; and the Brauer Library dedicated to Mathematics, Statistical Science, Computer Science, and Physics. The local Research Triangle region, which includes RTI International, FHI 360, Duke University, and North Carolina State University, is also an important resource of statistical and biomedical research. Major equipment and resources in the UNC Department of Biostatistics include:

General Computing

The Department of Biostatistics performs a large quantity of statistical computing and simulation for research and teaching purposes. As a result, the computing landscape in the Department is varied and wide. Currently the Department consists of several different work units with varied needs and varied equipment. The equipment is spread over various networks and two computer labs that are interconnected via the campus backbone. SuSE Linux Enterprise Server (SLES), Red Hat Enterprise Server (RHEL), Windows 2008 R2, Windows 2012 R2, Windows 7 are the supported Operating Systems with a wide range of computing equipment from IBM, HP, Dell and other manufacturers.

The Department of Biostatistics’ network presently operates from 113 Physical Servers, 38 Virtual Servers, 70 Thin-clients, 107 Physical Desktops and 45 laptops: 116 SLES11 64-bit Servers, 21 Windows servers and 5 VMWare ESXi servers for hosting virtual machines. Two of the SLES 11 Virtual servers host the department’s logon nodes in our high performance computing cluster (HPC), 1 SLES 11 server hosting the SLURM Master which schedules and dispatches all jobs submitted to the computing nodes. 98 of the physical servers are dedicated as computing nodes, running SLES 11 64-bit, and are responsible for processing jobs that are submitted from the logon nodes. 1 server is dedicated for very high memory intensive jobs and has 256GB of memory available. There are 5 ESXi servers are used to host 38 Virtual Machine and each Virtual Machine is dedicated for a specific service such as, print services, LDAP caching, SYSLOG, Splunk, DNS, DHCP, Proxy, Network Access Control, Database, Web, and thin client management. All servers with the exception of the virtual machines are running RAID-1 disk arrays. All Departmental data is hosted on a flexible and redundant storage cluster from the manufacturer EMC Isilon and currently has 172TB of space available for data storage. Data from the Isilon cluster is mirrored 24/7 to a separate physical location for backup and Disaster Recovery purposes. Monday through Friday, our secondary storage takes a snapshot of the Departmental data and is used for backup purposes. Snapshots are stored for 60 days then removed from the system. All desktops in the department are connected to the network using 1000MB Ethernet, all laptops are connected to the network via 1000MB networking or 54Mb wireless. The local network consists of several switches connected by fiber optic cable. Each of the local area network connects to the campus backbone and maintains up too 20Gb transfer rates.

The Department of Biostatistics owns and maintains about 441 computers and networked peripherals. This includes IBM-compatible machines, laptops, network printers, network scanners, and other network devices. By offering application serving, web and print services our computer facilities continue to meet the department’s needs for research, training, and administration.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Servers** | **Workstations** | **Laptops** | **Printers** | **Thin Clients** |
| On-site | 151 | 107 | 0 | 38 | 70 |
| Off Campus | 0 | 19 | 45 | 11 | 0 |
| Total | 151 | 126 | 45 | 49 | 70 |
| **Total Devices: 478** | | | | | |

The Department of Biostatistics currently uses a very wide range of statistical, imaging, GIS and genetics applications used to analyze data. The list of applications consists of Affymetrix Powertools, bedtools, bioprospector, bismark, boost, bowtie, brat, bsmap, casava, ccseg, cufflinks, dtitk, dti tract, epd, fastphase, fsl, freesurfer, gambit, gcta, hammer, hapgen, hapstat, hdf, hyperlasso, ibeat, igvtools, itk, isolasso, jags, mach, maq, methylseqpipeline, minimac, penncnv, plink, plinkseq, Qt, Revolution R, rrbsmap, samtools, statscan, shapeworks, slicer, spharm, tabix, tophat, vcftools, VisualAnalyticTkt, wsdl, Arcview GIS, CART, DBMS Copy, Dchip, Fiber Tracking, Fiber Viewer, FSL, Head Circumference, IRIS, ITK-SNAP, GenePix, JMP, Map Maker, Map Manager QTX, Mathematica, Matlab, MRIcro, MRI Watcher, nQuery, Qtlcart, R, S.A.G.E., SAS, ScanAlyze, Slicer, SNAP, Splus, Stata, StatXact, Sudaan, Treeview, Valmet, and Winbugs.

Security for Biostatistics Computing

The servers and clients are separated in to 3 firewalled networks ( zones ) which are managed and monitored by central campus. In addition to the network based firewalls, all hosts have strict host based firewalls to further reduce possibilities of internal or external network based attacks. Campus logs all network traffic that passes through the firewall and notifies of any anomalous events. All Windows and Linux server logs are sent to our central loghosts which monitor and alerts for any type of anomalous activities. All servers undergo weekly Qualys scans and any level 4 or 5 vulnerabilities are remediated within 30 days. Microsoft Windows laptops use full disk encryption and all MAC laptops use home directory encryption. Both types of encryption meet the FIPS 140-2 standard for auditing purposes.

Patch and application management for all windows desktops and servers and managed centrally from Microsoft’s System Center Configuration Manager (SCCM). Reports from SCCM are generated monthly and reviewed for any failed installations and action is taken within a reasonable time to remediate any issues. Patch and application management for all linux servers are centrally managed through custom designed scripts for ease of administration.

*Campus Research Computing:* ITS Research Computing provides resources to support and promote computational research at UNC-Chapel Hill. High performance and high throughput computing resources include a 704 node (8064 core) Dell Linux cluster, Kill Devil, with QDR Infiniband interconnect and a minimum of 4 GB memory per core; a smaller 2300-core HP Linux cluster with QDR Infiniband interconnect and at least 6 GB of memory per core; and two 32-core servers with one terabyte of memory each to accommodate codes that require extremely large amounts of RAM. The Kill Devil cluster also includes 64 NVidia Tesla GPUs (M2070).

Storage for research data accessed on the above systems includes more than 2.5 petabytes of disk, comprising locally attached disks; network-attached shared scratch space; and network-attached shared file systems. Supported file systems include Lustre, Isilon OneFS, GPFS and NFS. An archival mass storage system, with a current capacity of more than 700 TB, is also available. Infrastructure components are housed in an 11,000 square foot data center with more than 2 megawatts of power 800 tons of cooling available. The data center is staffed and monitored 7x24x365.

Centrally provided and managed software applications include a variety of Fortran and C compilers; Gaussian, Amber, Insight II, and dozens of other open-source packages commonly used in the biosciences and physical sciences. In addition, standard mathematical and statistical software such as Mathematica, Matlab, Stata, R, and SAS are available, as are GIS applications, including ArcGIS, and visualization/image processing packages such as ENVI and IDL. Virtual computing software developed at North Carolina State University has been deployed on many nodes of the general purpose compute cluster to allow researchers to run Windows or Linux-based applications easily on central computing systems. Computational scientists on staff in Research Computing are available for consultation on use of systems, applications, modeling, analysis and code optimization.

An Oracle database server and campus-wide Oracle site license can be used to provide backend database capabilities for applications and projects. In addition to hosting research databases, Research Computing’s database administrator can be leveraged by projects to ensure professional management of databases, freeing up researchers to focus on their actual research. Expert assistance is also available for performance analysis, code porting and code parallelization.

Condor-based grid capabilities have been deployed centrally and in various campus units, and Research Computing staff can also facilitate access to Open Science Grid and XSEDE national resources. In addition, Research Computing provides and supports a CentOS-based Linux image (Tar Heel Linux), customized for use with the UNC technology infrastructure, and a software application repository to enable rapid deployment of distributed Linux workstations on campus. Codes developed and tested in these environments can easily scale up to the central compute clusters.

Computing services in Research Computing are largely based on an underlying common campus architecture that includes Kerberos authentication, role-based authorization and shared, centrally installed and managed software applications when practical. More than two hundred software packages and utilities are offered for use on the central systems. By relying on ITS to maintain the hardware, security environment and software builds of computing systems, researchers are free to devote their time to science and research rather than to system administration.

Office

The Department of Biostatistics occupies a total of 32,099 square feet in three locations. One of these locations is on campus in the McGavran-Greenberg Building, one is adjoining the campus (Collaborative Studies Coordinating Center), and the third is approximately one-half mile from the campus (Carolina Survey Research Laboratory). Of this, about 25% is used for departmental administration and service facilities. The remainder is assigned to faculty, students, and research staff for education and research activities.

Meeting Space

The Department of Biostatistics has a state of the art conference room located in McGavran-Greenberg Hall which seats 20 people. It has a drop-down projector, wireless capabilities and conference calling facilities. The Gillings School of Global Public Health (GSGPH) has 14 conference rooms available to all GSGPH departments furnished with drop-down projectors and internet connections. Also available is the Blue Cross and Blue Shield of North Carolina Foundation Auditorium, the largest meeting space in the new Michael Hooker research center (physically connected to McGavran-Greenberg Hall). It has drop-down projectors and 104 built-in desks equipped with power and Internet connections. Five built-in cameras add video-conferencing capabilities for the School’s flourishing distance education and outreach programs.

*Updated: 9/22/16*