

## NEW YORK UNIVERSITY



“Our center needed a cluster that was flexible and heterogeneous but at the same time could be easily managed. Silicon Mechanics provided multiple options on how to achieve our requirements.”

*Alexander Statnikov Ph.D.  
Adjunct Associate Professor, CHIBI*

### Challenge

The NYU Center for Health Informatics and Bioinformatics' (CHIBI) HPC (high-performance cluster) was destroyed by flooding during Hurricane Sandy. Fortunately, the laboratory's director was able to unrack the data storage hard drives and rescue them from the advancing flood waters.

CHIBI needed to replace its computing capability quickly to respond to data requests from the many different researchers working on projects that rely on high-performance computing.

The challenge was to quickly configure and install a solution that had sufficient compute resources, and of the correct kind, to enable researchers to make discoveries while meeting the center's power, cooling and data center space requirements. This challenge was complicated by the fact that the center includes three very different use cases:

- Researchers assembling a new genome require compute intensive resources with huge amounts of memory and many memory nodes to be able to upload DNA.
- Those studying disease transmission and developing statistical models do not need as much memory, but may require a large amount of processor cores, possibly up to thousands of them.
- Researchers collecting data coming out of DNA sequencers do not require large amounts of memory or processor cores but rather cores that connect to the data, so I/O (input/output) throughput is more important.

consultation to understand CHIBI's hardware, software, cluster management and data storage requirements. Silicon Mechanics also provided on-site installation, racking, cabling, training and ongoing support.

The collaboration resulted in a swift deployment of the new HPC cluster, enabling the researchers to quickly access data and continue their research.

CHIBI focuses on research and education in bioinformatics and medical informatics, in particular on nine key areas with the most potential for novel discoveries and improved care including:

- Informatics and computational analytics for molecular signatures and personalized medicine.
- Computational causal discovery methods in biomedicine.

### Solution

NYU worked closely with Silicon Mechanics to configure a new HPC. The work included a thorough

- Next generation sequencing informatics.
- Cutting-edge educational informatics methods and systems.
- Next generation biomedical information retrieval and scientometrics.
- Innovative research design and analysis services.
- Assay-specific informatics for a variety of genomic and proteomic platforms.
- Advanced services and systems for multimodal data storage and management.
- Advanced enterprise-wide data mining methods, software and services.

## Results

CHIBI's work includes mathematical modeling on disease transmission, collection of data from a variety of instruments and genomics research, including the study of how genomic mutations can contribute to causing cancer.

One key project being undertaken is identifying whether some of the modern molecular high-throughput assay technologies (microarray gene expression, miRNA expression, proteomics, DNA methylation) are more useful than others and finding useful ways to combine multimodal data. Each modality was analyzed separately and then integrative multimodal classification techniques were applied. The Silicon Mechanics powered HPC cluster reduces the time needed to conduct statistical analyses, ensuring that findings are reproducible.

In addition, CHIBI is helping to set standards and best practices for text categorization through their study of machine learning algorithms. By studying which algorithms work the best, researchers can determine if a document is relevant to a query. The results provide useful insights to the research community. The researchers working on this study evaluated hundreds of text categorization examples resulting in extremely large data sets. Their findings ultimately increased the number of text categorization options and methods previously available to the research community.

“The powerful new computing resources in the updated HPC enable our researchers to tackle difficult biomedical problems, as well as help CHIBI attract funding and talented researchers to join the institution to extend its research.”

*Efstathios Efstathiadis Ph.D.  
Technical Director,  
High Performance Computing Facility,  
CHIBI*

## HPC Configuration

- 2 head nodes, 2 x 2U = 4U
- 16 compute nodes, 16 x 2U = 32U
- 1 high memory node, 1 x 2U = 2U
- 5 GPU nodes, 5 x 1U = 5U
- 1 storage server, 81TB, 1 x 4U = 4U
- 6 service nodes, 6 x 2U = 12U
- 1 processing node, 1 x 2U = 2U
- 2 image annotation and distribution servers, 2 x 2U = 4U
- Bright Cluster Manager®
- Cluster integration software
- On-site integration of all systems, network and cabling, training
- Software and hardware support

