Cornell University
Center for Advanced Computing

CAC
we enable your success

CAC and the North Shore LIJ Health System

The Feinstein Biorepository: Designing an Efficient Data Management System for Research

How do you effectively manage and mine large-scale databases of clinical and laboratory data that will include billions of SNP genotypes?

Finding the Answer

CAC systems and database consultants helped design and create a new high-performance computing center and informatics system for the Biorepository at The Feinstein Institute for Medical Research.

The Feinstein Biorepository

Located at the largest public healthcare system in New York State, the Biorepository at the Feinstein Institute for Medical Research has grown to store hundreds of thousands of human sample of different types, such as serum, plasma, DNA, cells, tissues and tumors, along with extensive amounts of associated data, to support many large scientific studies.



Both control and disease-affected samples are collected and managed along with clinical, laboratory and bioinformatics data.

Improved Research

Research Metrics

- Scalability: High-performance computing systems must scale to handle billions of SNPs, plus clinical data
- Performance: Architect a system with CAC consultants to insure proper memory, processor, and network

Research Challenge

One segment of the sample analysis that has grown dramatically is the identification of single nucleotide polymorphisms, or SNPs. SNPs are DNA sequence variations that occur when a single nucleotide in a genome sequence is altered. SNPs make up about 90% of all human genetic variation and scientists believe that SNPs may predispose people to a disease or influence their response to a drug. Researchers at The Feinstein Institute are generating approximately 8 to 10 million SNP genotypes each day, and they anticipate accumulating billions of SNPs.

“The difficulties in managing and manipulating these very large datasets required the creation of a new data center capable of high-performance data management,” said Robert Lundsten, Biorepository Director. “Management of research-subject annotation is also quickly becoming a high-performance computing issue,” he added.

Solution

CAC systems and database consultants helped The Feinstein Institute architect a data management solution that would meet their current needs and scale to handle the rapidly growing SNP data.

A symmetrical multi-processor (SMP) Unisys system was deployed. The platform was designed to run SQL Server. Data is stored directly through four host bus adapters to an EMC CLARiiON RAID disk array. The computer center also has an assortment of in-house life science applications running on Dell PowerEdge servers and Dell PowerVault disk arrays.

The Client

North Shore – LIJ Health System

- New York State’s largest health care network with over 7,000 doctors, over 10,000 nurses
- Headquarters in Manhasset, NY
- Among top 6% of all institutions nationally receiving NIH funding
- Goal is to understand biological processes that underlie various diseases and translate this knowledge into new tools for diagnosis and treatment

The Collaborative Relationship

CAC's expertise in high-performance data management solutions was a perfect match for The Feinstein Institute's data requirements.

"Creating an efficient data management environment is the first step in developing an effective data mining environment. CAC's experience in data management design was very helpful. They know how to design systems and databases that optimized performance in SQL environments."

Robert Lundsten
Biorepository Director
The Feinstein Institute for Medical Research
North Shore LIG Health System