

## **Bioinformatics Cluster at ITTC**

### **Background**

A computing facility (including computing cluster, storage facility, backup system and associated software) facilitates the development, testing and deployment of new computing technologies in support of research on a variety of life sciences problems. The ITTC facility provides for research into new algorithms and methods for genomics (including analysis of microarray data), proteomics, molecular dynamics, molecular docking and analysis of magneto-encephalography (MEG) data. The facility also supports various public domain software and academic licensed software to serve needs of our research and development community. The close collaboration between chemists, biologists, mathematicians, and computer scientists using this facility results in the creation of new computing and information technologies and also directly leads to increasing our understanding of biological systems.

### **History**

The Bioinformatics Cluster at ITTC was built as a 16-node dual-processor Beowulf cluster connected with 25 terabytes of storage, constructed in 2005. The system was implemented to provide a high-performance computing environment, coupled with high-availability storage and backup solution. The system was expanded in 2006 with the addition of 64 quad-core machines and again in 2009 with 48 eight-core machines.

### **Cluster Size and Users**

ITTC's open access scientific computing system received an additional 384 cores through an NSF grant, taking it to a total of 768 cores (processors) with 176 nodes.

By contrast, the KUMC computing cluster has 32 nodes and 64 clusters and the Bioinformatics Cluster at the Center for Bioinformatics on the KU campus has 175 nodes and 646 cores.

The cluster is owned by ITTC, and is managed by the ITTC IT manager. It is open to ITTC faculty, students and staff, as well as any academic researcher approved by the ITTC Director. Fee-based access is available to non-academic users. For example, researchers at Hill's Pet Food and Porter McGuffey have obtained fee-based access to the system.

## Current Statistics

<b>Bioinformatics Cluster at the Information and Telecommunication Technology Center Statistics</b>	
Number of Nodes	176
Processors Total	768
Theoretical Peak Performance	6600 GFlops
Total Peak Performance	4667 GFlops
Total Memory	1408 GB
Total Disk	22528 GB
Operating System	Linux
Interconnect Technology	Gigabit Ethernet (x2), InfiniBand on sel
Communication Software	MPICH2, MVAPICH, Ope
Main Application Area	Bioinformatics
Year Installed	2005
<b>Each Node</b>	
Type of Processor	Intel Xeon EM64T
Processors per Node	2x64 - 4x64 - 8x48
Frequency	2x3.2 Ghz 4x2.8 Ghz 8x2.833GHz
Node Peak Performance	2x7GFlops 4x14GFlops 8x75GFlops
Memory per Node	2x4096 MB- 4x6144 MB- 8x16384 MB
Disk per Node	128x146 GB , 48x80 G
<b>Filesystem Node</b>	
Type of Processor	Intel Xeon EM64T
Processors	2
Frequency	3200 Mhz
Total Disk	146 GB
Total Disk	4096 MB
<b>Filesystem (SAN)</b>	
Dual RAID Controllers	
17T (10000 GB) SATA RAID 5 (6+1)	
15T (15000 GB) Fibre Channel RAID 5 (7+1)	