SGI™ High-Throughput Computing (HTC) Environment for Bioinformatics

Features

- High-performance, cost-effective bioinformatics data-processing solution
- A high-throughput SGI IRIX cluster distribution model for BLAST, Clustal W, FASTA, and HMMER
- Optimized for ultrahigh-density servers (2–4 MIPS CPUs in a 3.5-inch rack-mountable chassis)
- Advanced data and file management capabilities for rapid access to large data sets
- Cluster administration and management tools

Benefits

- Save time and money with a proven, supported, high-performance production computing environment
- System scalability simplifies expansion due to growth in data volume and processing power
- Flexible throughput model permits future expansion to include additional applications

A High-Performance Bioinformatics Environment

The SGI HTC environment for bioinformatics brings together SGI expertise in bioinformatics applications and SGI™ Origin™ 300, a scalable server designed for cost-effective large computing environments. The SGI HTC environment for bioinformatics delivers outstanding computing power at a substantially lower price, offering the best combination of performance and scalability for bioinformatics applications.

HTC wrappers were developed by SGI applications engineers to maximize throughput and utilization of an SGI Origin 300 cluster for the bioinformatics applications BLAST, FASTA, Clustal W, and HMMER. This solution can be easily integrated into production computing environments in a manner that is transparent to the end user. The HTC environment is designed to act as a drop-in replacement for the traditional commands so that no or very small changes to scripts or existing workflows should be needed to incorporate it into a bioinformatics computational production environment.

Designed to Grow

The scalability of the HTC applications and the simplicity of the SGI HTC environment for bioinformatics allow for maximum system utilization and for uncomplicated system expansion. Furthermore, the architecture of the HTC environment is extensible to support the addition of future applications that benefit from improved distribution models for high-throughput calculations.

Total Cost of Ownership

When considering a solution for a cost-effective bioinformatics data processing environment, many customers look to the low-cost 32-bit Intel® architecture hardware coupled with the Linux® operating system. Although such solutions can offer low upfront costs associated with the initial hardware purchase, achieving a production-level Linux cluster typically requires significant consulting and administrative fees. Based on interviews with customers who have built Linux bioinformatics solutions, we estimate that customers would pay up to 70% more over three years to purchase and operate a Linux cluster of 32 1 GHz Pentium® III processors than they would pay for a computationally equivalent SGI Origin 300 cluster of 12 (3x4 CPUs) 500 MHz MIPS® processors.

Storage and File Management

Utilizing the SGI™ Clustered XFS™ (CXFS™) file management system and an SGI storage system such as SGI™ TP900, the SGI HTC environment for bioinformatics can provide a highly efficient centralized storage solution for accessing and managing databases. CXFS provides a single filesystem view that spans a cluster and that has performance characteristics similar to those of XFS, the industry-leading SGI filesystem. The SGI TP900 storage system is a modular, high-performance JBOD storage device that connects easily as disk modules to the integrated SCSI port in the SGI Origin 300 server. With a choice of one or two Ultra160 SCSI channels and the latest, state-of-the-art SCSI 10,000 RPM and 15,000 RPM drives, the TP900 can be optimized to meet customers' expectations.

SGI High-Throughput Computing Environment for Bioinformatics Technical Specifications

System Configurations
- SGI Origin 300 servers clustered together with standard Ethernet
- Optional cluster connections with Myrinet or SGI NUMAlink™
- Cluster size: from 4 processors with Ethernet and Myrinet connections
- Shared memory size: 4 processors to 32 processors with SGI NUMAlink connections

SGI Total Performance 900 (TP900) Storage System
- Eight Ultra160 SCSI JBOD drives
- Up to 584GB in 3.5"

Operating System Support
- IRIX® version 6.5.x
- Fully supported by SGI Global Services

HTC-BLAST, HTC-FASTA, HTC-Clustal W, HTC-HMMER
- High-throughput cluster distribution model for NCBI-BLAST, WU-BLAST, FASTA, Clustal W, and HMMER programs; other applications available upon request
- Modifications optimize the high-throughput processing of multiple bioinformatics jobs and do not affect the functionality or results of the original applications

Genomics Database Management
- High-performance CXFS file serving

SGI Origin 300 Server
- 2U (3.5") chassis
- 2 or 4 MIPS R14000™ processors
- Up to 4GB of SDRAM memory