BioTeam Convergence Lab

Established in January 2016, BioTeam launched this lab with contributions from Silicon Mechanics to enable future scientific advancements at Texas Advanced Computing Center. The concept of this cooperative environment, led by BioTeam's VP and General Manager of Consulting Services, Ari Berman, was initiated to develop innovative architectures that are designed to push the boundaries of research computing. Together with the hardware expertise provided by Silicon Mechanics and the in depth knowledge of scientific technologies from BioTeam, both firms are committed to developing the lab.
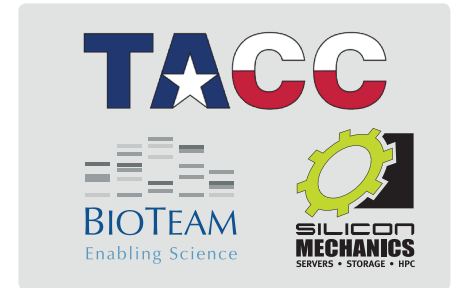
Cluster Configuration

The cluster is racked to standard procedures in a 42U rack. Network connectivity is provided by BioTeam--configured Juniper hardware in the form of and MX Router and EX switch. All inter-node connectivity is through copper 10G, with copper 1G lines for administration.

BioTeam configured the JBOD with ZFS for a total of 44TB usable storage. The compute servers currently have Centos 7 installed and various bioinformatics programs. All devices have been stable for the 1 year they have been in operation and have demonstrated no obvious issues.

Current Cluster Use Cases

Remote backup for home directories for Government HPC cluster
* The ZFS formatted JBOD was utilized by BioTeam to serve as a mirror for the home directories of over 200 researchers from a remote HPC cluster.

Network Address Translation, SSH, and VPN for access to the Government compute cluster
* The JBOD controller was used to handle various networking needs for BioTeam and a government agency, including Network Address Translation and remote connectivity.

Data ingest service for data transfers to remote Government HPC storage
* The JBOD and controller node were used to transfer large (200GB--1TB) life sciences datasets to a remote compute cluster primary storage. Due to current agency network limitations, the Convergence Lab has served as a high speed data transfer service.

Benchmarking for bioinformatics workflows
* The compute servers were used for indexing the hg381 version of the human genome for testing disk performance . The Burrows-Wheeler-Aligner (BWA)[2] application was used to create an index of the human genome. A single index job was run as well as four simultaneous indexing jobs writing to a network disk. In these cases, two different vendor Network Attached Devices were evaluated. Other workflows are ongoing.

Silicon Mechanics Hardware Overview

(2) Compute Nodes Rackform Rackform U638	• 2 x Intel Xeon E5-2650v3, 2.3 GHz (10-Core, HT, 25MB Cache, 105W) 256 GB RAM
(2) GPU Nodes Rackform Rackform U638	• 2 x Intel Xeon E5-2650v3, 2.3 GHz (10-Core, HT, 25MB Cache, 105W) 256 GB RAM
	• 2 x NVIDIA Tesla K80m GPU, 24GB GDDR5
(1) 4U JBOD 946ED-R2KJBOD	• 46 X 2TB 3.5 HDD
(1) JBOD Controller (Software Defined Storage Node) Rackfrom U638	• 2x Intel Xeon E5-2643v3, 3.4 GHz (6C, HT, 20MB Cache, 135W) 256 GB RAM

1. Human Reference Genome; http://hgdownload.cse.ucsc.edu/downloads.html#human

2. Li H. and Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25:1754-60 [PMID:19451168] http://bio--bwa.sourceforge.net/

Expert included.