

Siberian supercomputer center as a service for bioinformatics research

I. Chernykh, B. Glinskiy, N. Kuchin, S. Lomakin

Institute of Computational Mathematics and Mathematical Geophysics SB RAS, Novosibirsk, Russia

* e-mail: chernykh@ssd.ssc.ru

Key words: high performance computing, GPU, bioinformatics and life sciences

Introduction: Sequencing and protein docking are very compute-intensive tasks that see a large performance benefit by using the latest HPC hardware. At this moment there are a lot of bioinformatics codes which are optimized for the latest Intel HPC hardware [1]. Siberian Supercomputer Center (SSCC) has new HPC cluster with total peak performance ~91 TFlops. This system is well designed for bioinformatics researches due to the using Intel Xeon Phi (KNL architecture) CPUs as well as Intel Optane technology for extending memory size on Intel CPU node.

Siberian Supercomputer Center resources: SSCC offers computer resources for bioinformatics researches to its users. Our main system NKS-1P consists of 40 Intel Xeon E5-2697v4 (2.6 GHz, 16 cores) and 16 Intel Xeon Phi 7290 KNL (1.5 GHz, 72 cores, 16 GB MCDRAM) CPUs. Intel Xeon E5-2697v4 CPU nodes have 128 GB DRAM, Intel Xeon Phi 7290 nodes have 96 GB DRAM. For bioinformatics problems, we have 2x 375GB Intel Optane memory which is working as IMDT on Broadwell node. HPC nodes and 200TB Intel Lustre PFS are working on Intel OmniPath 100 Gb/s interconnect. We also have the supercomputer with a hybrid architecture and consists of NKS-30T (platform BL2h220c hp) system with 576 Intel Xeon processors E5450/E5540/X5670 (2688 cores) and hybrid cluster that based on 40 servers HP SL390s G7 (80x CPU X5670 – 480 cores) with 3x NVidia Tesla M2090 GPU on each node. All cluster nodes are connected via Infiniband QDR network interface. Cluster file system IBRIX (4 servers, 32 TB of available disk space) is also connected by Infiniband interface for NKS-30T. The newest part of SSCC resources is based on [2]. This architecture is well suited for open source packages like MUMmerGPU: High-through DNA sequence alignment using GPUs [3], Parallel-META: a GPU- and multi-core-CPU-based open-source pipeline for metagenomic data analysis, which enabled the efficient and parallel analysis of multiple metagenomic datasets [4], and Molecular Dynamics packages like GROMACS [5], LAMMPS [6]. All these resources are available to all organizations that are operated by the Federal Agency of Scientific Organizations Russia.

References

1. <https://www.intel.com/content/www/us/en/healthcare-it/solutions/genomicscode.html>
2. <http://www.ssc.ru/iccmmg.nsc.ru/hardware.html>
3. <https://sourceforge.net/projects/mummergpu/>
4. Su X., Xu J., Ning K. (2011) Parallel-META: A high-performance computational pipeline for metagenomic data analysis. 2011 IEEE International Conference on Systems Biology (ISB), Zhuhai. P. 173-178.
5. <http://www.gromacs.org/>
6. <http://lammms.sandia.gov/>
7. Kzantsev F. et al. (2008) Proc. of the 6th International Conference on BGRS. P. 113.
8. Likhosvai V. et al. (2001) Generalized chemokinetic method for gene network simulation. Mol Biol. 35:1072-1079.