

OOKAMI PROJECT APPLICATION

Date: August 09,2021

Project Title: Exploring Computer architectural requirements for single cell genomics applications

Usage:

- Testbed

Principal Investigator: Dirk Pleiter

- University/Company/Institute: Forschungszentrum Juelich
- Mailing address including country: Wilhelm-Johnen-Strasse, Juelich Supercomputing Centre, 52428 Juelich, Germany
- Phone number: +49 2461 61-9327
- Email: d.pleiter@fz-juelich.de

Names & Email of initial project users:

Dirk Pleiter, D.pleiter@fz-juelich.de Fatemeh Pouyan, F.pouyan@fz-juelich.de

Usage Description:

We have been exploring computer architecture requirements for single cell genomics applications. We focus on Kallisto¹ as a genomics application which develops a method introduced as near-optimal probabilistic RNA-seq quantification. So far, we have investigated different parameters and factors that impact on performance of Kallisto on diverse architectures including Arm-based platforms at the Juelich Supercomputing Centre, but due to the lack of access to the A64FX platform we have not been able to do much work on this platform.

Performance profiling and benchmarking on different architectures with respect to compute capabilities would allow us to discover parametric dependency between architecture components and the overall performance of Kallisto. Kallisto implements a hash table to match read k-mers to compatible transcripts. The hash table should be stored in memory. Each time a k-mer is looked up using the hash, several load operations are occurred due to the hash

¹<https://github.com/pachterlab/kallisto>

table is not collision-free. The runtime of Kallisto is influenced by looking up process. Thus, we foresee that the high-bandwidth unified memory of A64FX CPU would make a positive impact on the runtime of Kallisto. We already did primary benchmarking on the Fugaku supercomputer at the RIKEN research center in Japan and we observed that the back-end stall cycles on A64FX server was greater than the ThunderX2 server that resulted in lower execution time on the ThunderX2 Server [1]. Therefore, We would like to evaluate the performance of Kallisto on Ookami and investigate new ways of tuning hashing algorithms to improve the performance. Our long-term goal is to develop a performance model based on parameters of the application and performance metrics of the architecture in addition to investigate Arm-based solutions for future tenders. The model allows us to choose between different application optimisation and architecture options to make a prediction and knowledge about their influence on the overall performance of Kallisto.

Computational Resources:

- Total node hours per year: 1,000
- Size (nodes) and duration (hours) for a typical batch job: The job will need only one node. Typical job duration will be less than 1 hour.
- Disk space (home, project, scratch): 200 GBs will be sufficient

Personnel Resources (assistance in porting/tuning, or training for your users):

N/A

Required software:

GNU C/C++ compiler, Fujitsu C/C++ compiler, CMake, HDF5, PAPI, Autoconf if possible

If your research is supported by US federal agencies:

References:

[1] Bine Brank, Stepan Nassyr, Fatemeh Pouyan, Dirk Pleiter: Porting Applications to Arm-based Processors. CLUSTER 2020: 559-566