



CSSI Frameworks: Re-engineering Galaxy for Performance, Scalability and Energy Efficiency



PennState



Award #OAC-1931531

PI: Mahmut Taylan Kandemir *, Co-PIs: Chita R. Das *, Anton Nekrutenko Ψ , Paul Medvedev *
* Dept. of Computer Science and Engineering, Ψ Dept. of Biochemistry and Molecular Biology
Institution: The Pennsylvania State University, University Park

Motivation

- Galaxy currently runs on a large variety of high-performance computing (HPC) platforms including super computers.
- Bio-Toolchain execution in Galaxy take in order of days, thus hindering the pace of scientific discovery.
- GPUs and FPGAs have superior performance compared to CPU with up to 800x improvement in performance.
- Galaxy does not have support to use GPUs and accelerators like FPGAs.

Our Proposal

- Modernize** the Galaxy framework to utilize modern compute platforms such as GPUs and FPGAs.
- Rewrite** existing tools to support GPU and FPGA acceleration.
- Accelerator-Aware** Computation Mapping and Orchestration.
- Dynamic Resource **Scheduling** Based On Real-Time Feedback.
- Redesigning** Storage for Galaxy.

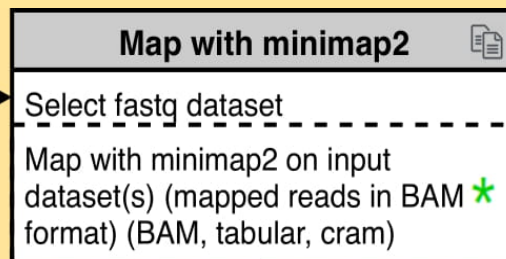
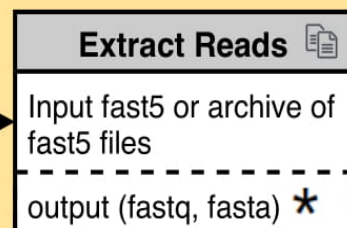
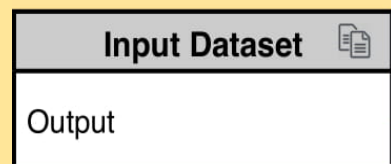
Completed Tasks

- Identified** existing GPU based tools.
- Integrated** RACON GPU based tool to Galaxy.
- Deployed** Galaxy on Kubernetes.

Planned Tasks

- Develop GPU based implementation for existing ML tools for RNN sequencing.
- Expose GPU metrics to Kubernetes for developing resource utilization-aware scheduling policies.
- Develop a Reinforcement-based Machine Learning scheduling framework.

Example Galaxy Bio-Toolchain



Galaxy Impact

300k page views 8915 publications > 7800 supported tools

Deployed in > 7 million Research funds

128 platforms