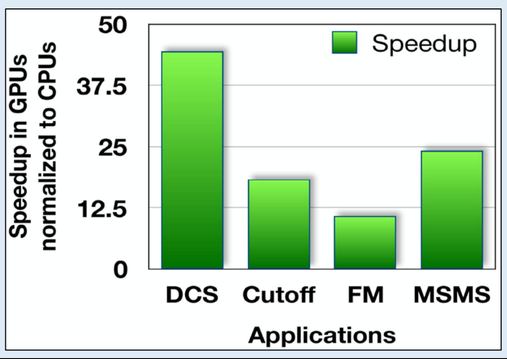
CSSI Frameworks: Re-engineering Galaxy for Performance, Scalability and Energy Efficiency

Award #: OAC-**1931531**

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

I. What is the Problem?

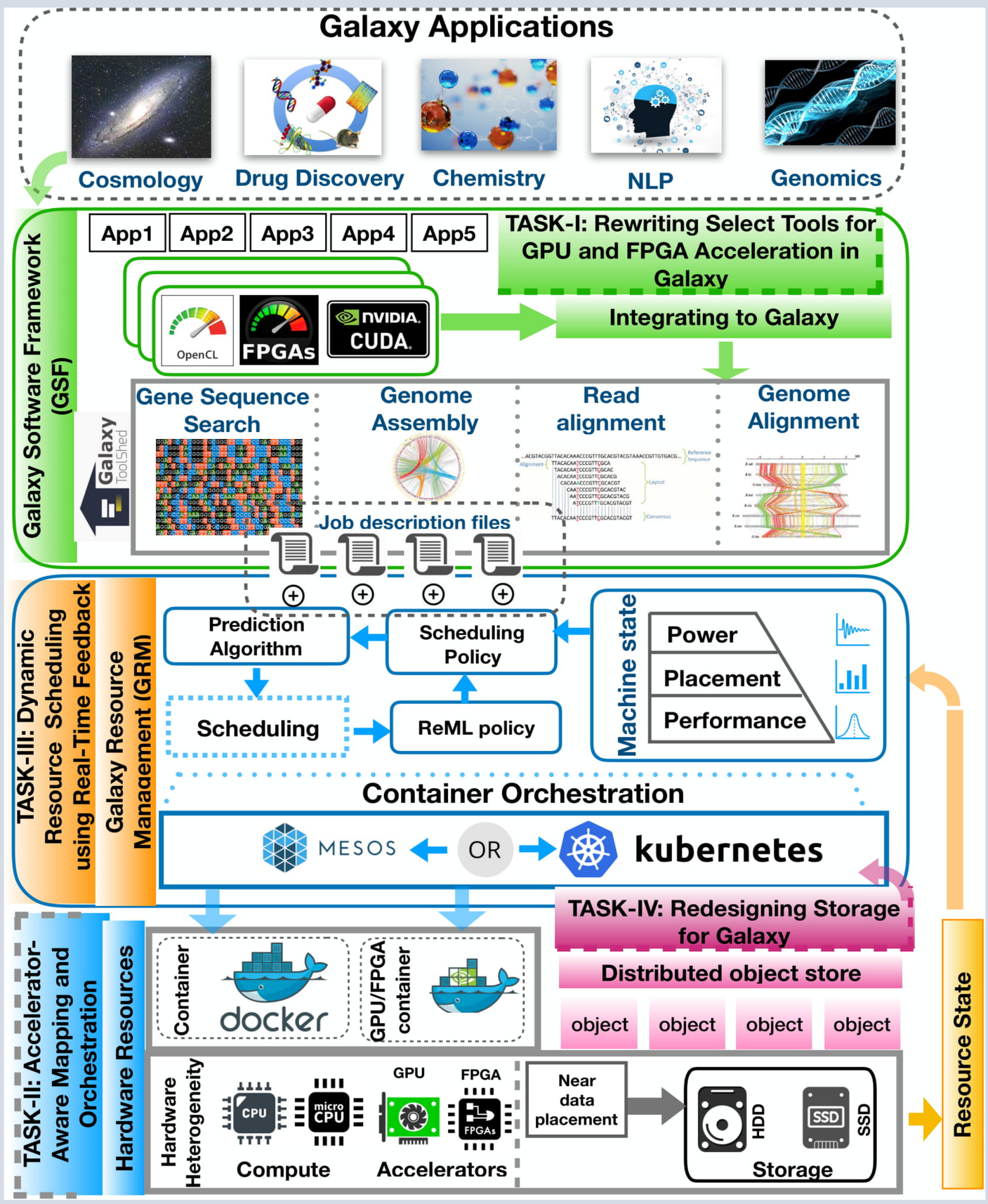
- Galaxy is an open source, web-based framework that is extensively used by more than 20,000 researchers world-wide in many areas.
- Current galaxy implementation does not support GPU and accelerators like FPGAs.
- There is no support for dynamic resource scheduling and management in Galaxy.



II. Our Objective

- **Re-engineer** the Galaxy framework to enable GPUs and accelerators like FPGAs as "first-class" compute engines.
- **Enlarge** the Galaxy community by bringing GPU and FPGA supported tools.
- Enable Galaxy tools to take better advantage of **emerging cluster scheduling** capabilities.
- Achieve significant improvements in performance **scalability** and **energy efficiency**.

III. Our Vision for Galaxy



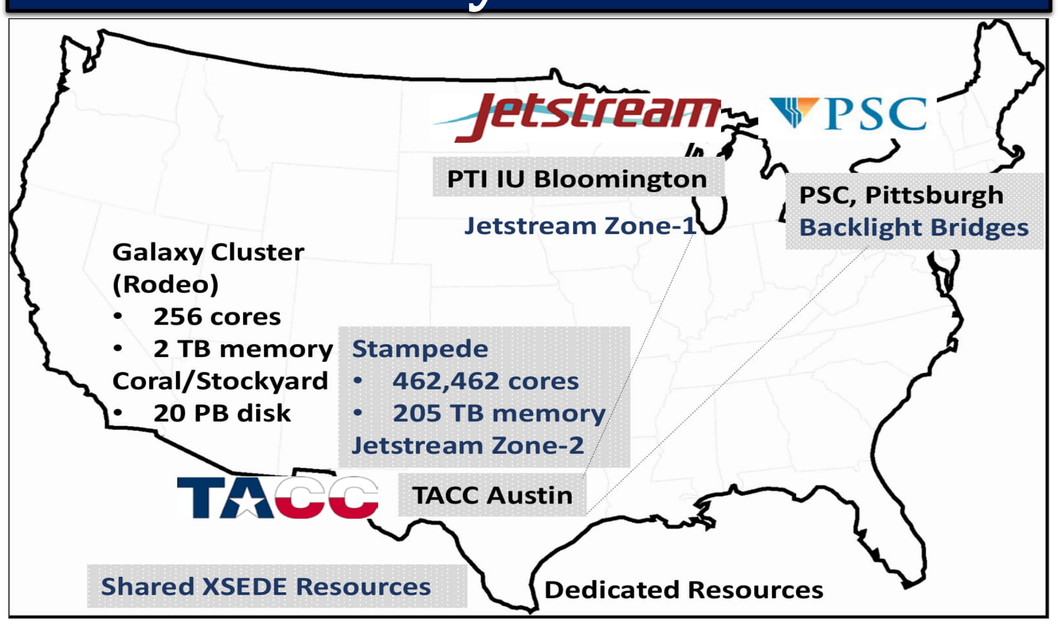
IV. Current Developments

- **Identified** existing GPU based tools such as Racon.
- **Integrated** a GPU based implementation of RACON, a genomic consensus tool to Galaxy Toolshed .
- **Integrated** a GPU-version of Smith-Waterman sequencing tool to Galaxy.
- **Deployed** Galaxy on a Kubernetes managed cluster.

V. Next Steps

- Develop GPU based implementation for existing ML tools for RNN sequencing, using Keras framework.
- Use GPU plugin in Kubernetes to enable GPU access for Galaxy pods.
- Expose GPU metrics to develop resource utilization-aware scheduling policies.
- Develop a Reinforcement-based Machine Learning scheduling framework.
- Develop a distributed data-store using Rook, to enable near data placement based scheduling policies.

Galaxy Centers



Galaxy Popularity

300k page views

8915 publications

> 7800 supported tools

10000+ online subscribers

Deployed in **128** platforms

> 7 million Research funds

8000+ twitter followers

Learn More

https://galaxyproject.org

Get in Touch: mtk2@psu.edu

Acknowledgments:
Galaxy project is also supported by NSF grant #1661497 and other NIH grants.