

Scalable Cyberinfrastructure for Artificial Intelligence and Likelihood-Free Inference



K.-P.-H. Anampa¹ J. Bonham² K. Cranmer⁴ (PI) B. Galewsky³ M. Hildreth¹ (PI)
D. S. Katz^{2,3} (co-PI) C. Kankel¹ I.-E. Morales⁴ H. Mueller⁴ (co-PI) M. Neubauer^{2,3} (PI)

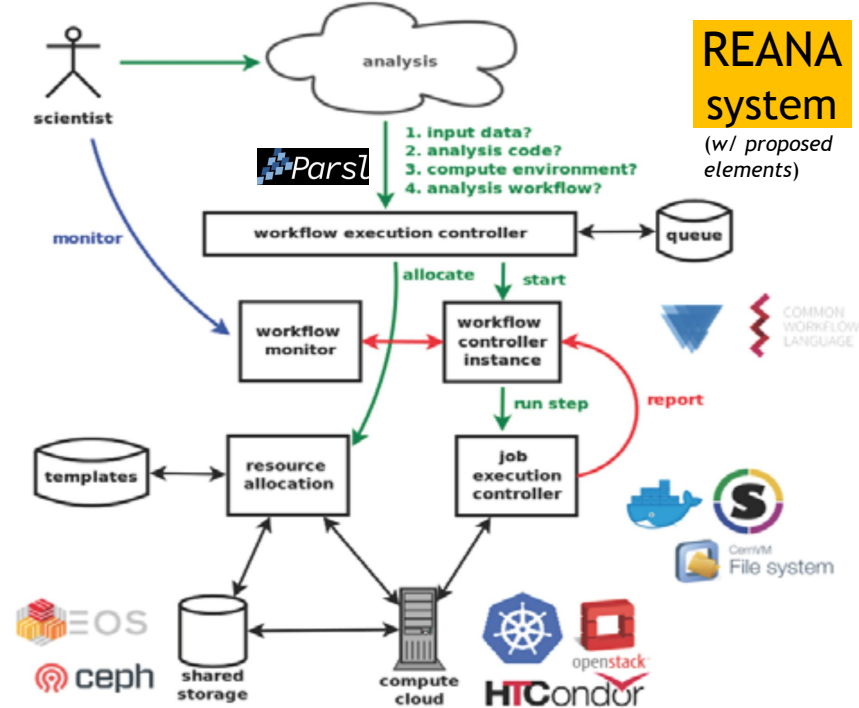
OAC: [1841456](#), [1841471](#), [1841448](#)

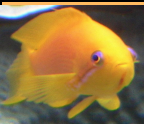
¹University of Notre Dame ²University of Illinois ³National Center for Supercomputing Applications ⁴New York University

scailfin.github.io

Main Goal

- To deploy **artificial intelligence** and **likelihood-free inference** methods and software using **scalable cyberinfrastructure** (CI) to be integrated into existing CI elements such as the *REANA system*, to **increase the discovery reach of data-intensive science**



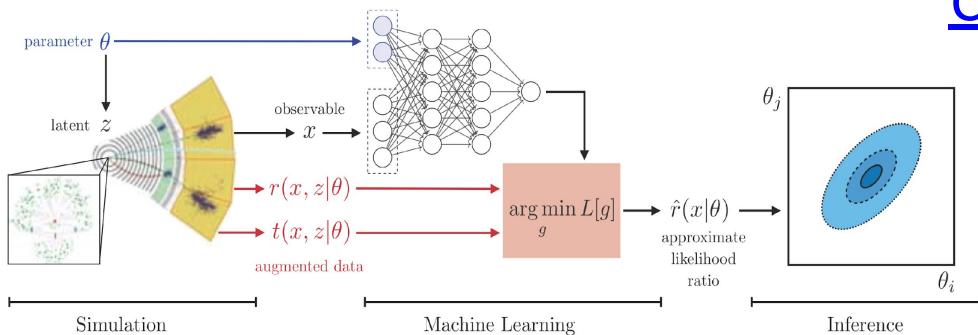


The SCALFIN Project



Likelihood-Free Inference

- Methods used to constrain the parameters of a model by finding the values which yield simulated data that closely resembles the observed data

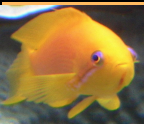


Catalyzing Convergent Research

- Current tools are limited by a lack of scalability for **data-intensive problems** with **computationally-intensive simulators**
- Tools will be designed to be **scalable** and **immediately deployable** on a **diverse set of computing resources**, including **HPCs**
- Integrating **common workflow languages** to drive an **optimization of machine learning elements** and to **orchestrate large scale workflows** **lowers the barrier-to-entry** for researchers from other science domains

Science Drivers

- Analysis of **data from the Large Hadron Collider** is the **primary science driver**, yet the technology is sufficiently generic to be **applicable to other scientific efforts**



SCAILFIN Project Activities



REANA Deployment and Application Development

- Established a shared REANA development cluster at NCSA
- REANA implementation of MadMiner workflow & ML-based top quark tagger comparisons
- Ongoing studies of Matrix Element Method approximations using deep neural networks

Parsl Integration

- *Parsl*: Annotate python functions to enable them to be run in parallel on laptops, OSG, supercomputers, clouds, or a combination without otherwise changing the original python program and developing capability to export workflow to CWL
- Have ported a REANA example workflow to Parsl

HPC Integration

- Using VC3 infrastructure to configure and set up edge service head node on a cluster at ND
- REANA runs on head node, submits jobs to HPC batch queue using HTCondor
- Jobs are now successfully submitted to worker nodes
 - “Hard problems” and new infrastructure ~finished; “simple issues” like file and executable transfer still to be solved for full chain to work
- Plan to move to integration and testing on the NSF Blue Waters Supercomputer at Illinois