Related NSF Projects

In support of HL-LHC R&D

Goal

- Leverage investments made by NSF in projects with overlapping R&D goals
- Benefit from contributions and experience in other contexts & communities
- Offers additional channels for collaboration and broader impacts - adoption of R&D products and solutions

scailfin.github.io

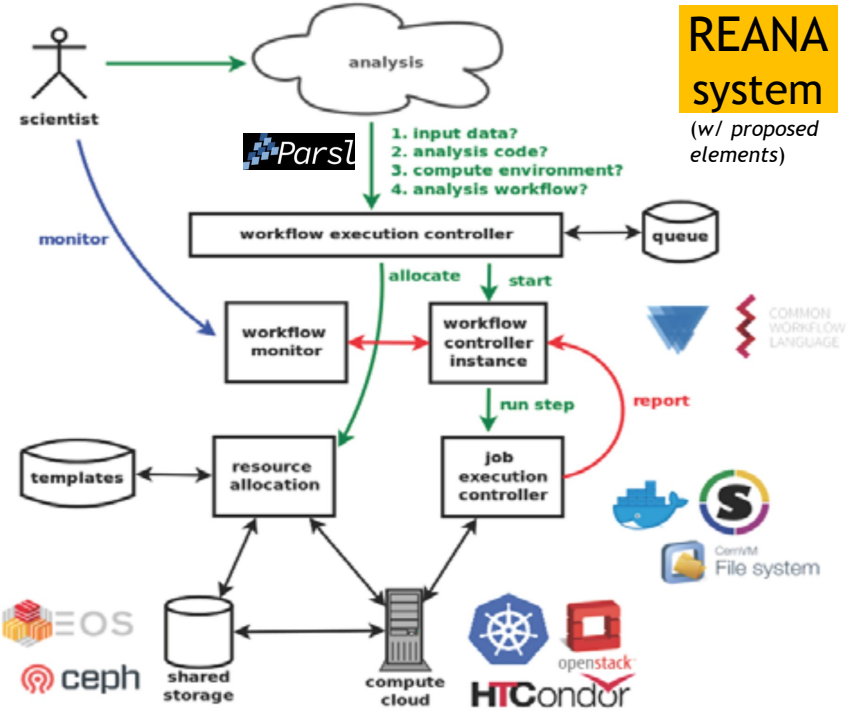
Scalable Cyberinfrastructure for Artificial Intelligence and Likelihood-Free Inference

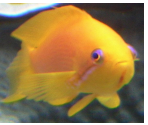
U. Illinois / NCSA, U. Notre Dame, NYU

NSF OAC-1841456, 1841471, 1841448

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***

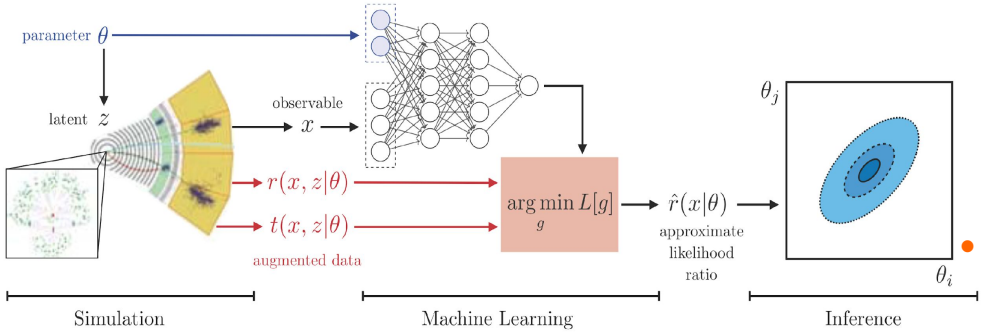




The SCAILFIN Project

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

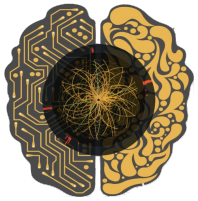


Science Drivers

- 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

Catalyzing Convergent Research

- Current tools are limited by a lack of scalability for data-intensive problems with computationally-intensive simulators
- Tools are being designed to be scalable and immediately deployable on a diverse set of computing resources, including ***HPCs***
  - REANA is now running on BlueWaters and NERSC (CORI)
- 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
  - We have ported a REANA example workflow to Parsl and have been investigating MLFlow



hls 4 ml

https://fastmachinelearning.org

Advancing Science with Accelerated Machine Learning

MIT/UIUC/UW NSF OAC-1934700, 1934757, 1934360

- Main Goal: Bring machine learning based accelerated computing into the scientific community by pushing the frontiers of deep learning at scale in order to **accelerate and enable new physics discovery** in the big data era
  - Developing machine learning (ML) based acceleration tools focusing on Field Programmable Gate Arrays (FPGAs) to accelerate low latency inference of ML algorithms
  - Building **FPGA demonstrators** to reduce overall computing latency for two benchmark experiments
    - **High level trigger event reconstruction** at the Large Hadron Collider (LHC)
    - **Gravitational wave identification** at the Laser Interferometer Gravitational-wave Observatory (LIGO)
- Developing open source tools readily shared with LHC, LIGO, and LSST
- Part of the NSF-OAC's Harnessing the Data Revolution (HDR) Big Idea activity

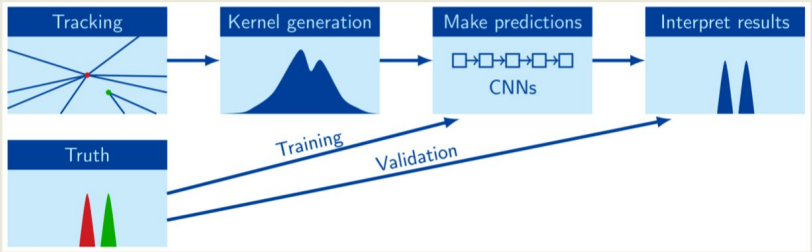
PV Finder NSF OAC-1740102 & OAC-1739772

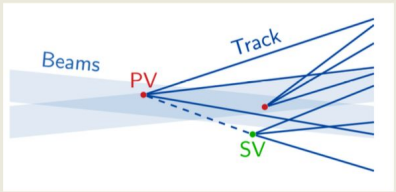
supported by NSF awards OAC-1740102 & OAC-1739772

G. Watts (UW/Seattle)

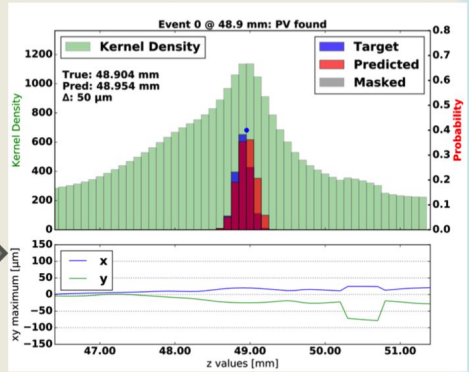
PV Finder

Primary Vertex finding @LHCb using CNN's





Can we do vertexing in a tracker like LHCb using ML?
Proof of Principle has been established
Next: training on large datasets & production in HLT



https://gitlab.cern.ch/LHCb-Reco-Dev/pv-finder

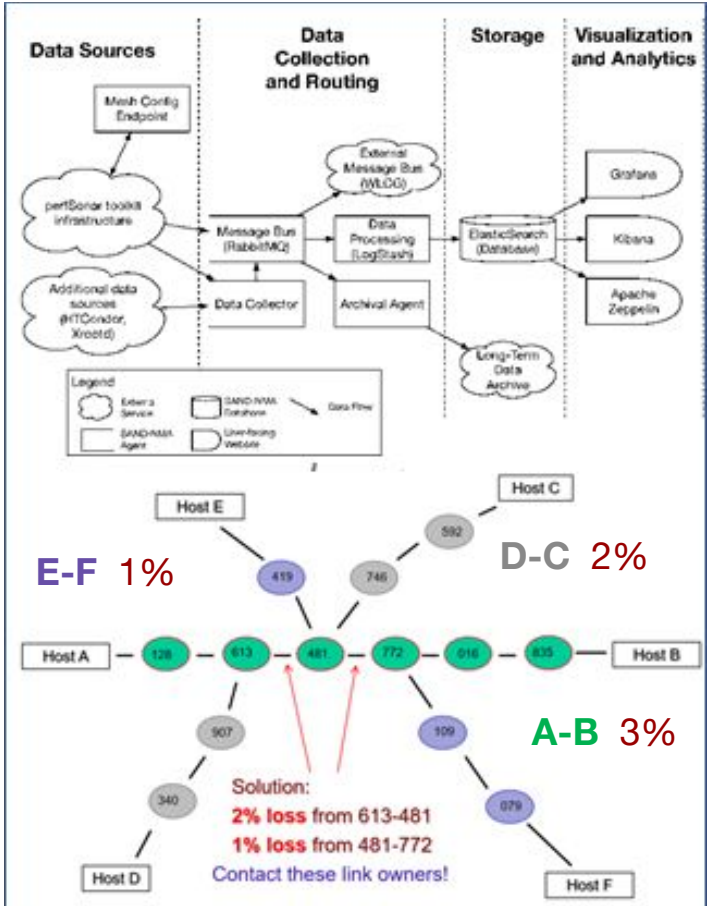
SAND Service Analysis and Network Diagnosis NSF OAC-1827116

Focuses on combining, visualizing, and analyzing disparate network monitoring **and services** logging data

Will **extend** and **augment** the **OSG networking** efforts with a **primary goal** of extracting useful insights and metrics from the wealth of network data being gathered from perfSONAR, FTS, R&E network flows and related network information from HTCondor and others.

Combine with LHC workload & data management analytics (shared ELK platform & analysis nodes)

Exploit network tomography we have by continuously measuring thousands of R&E network paths.



SLATE: Services Layer at the Edge NSF OAC-1724821

- Remotely manage edge services at sites by **expert teams** from **trusted organizations**
- Deploy updates more quickly & introduce new services more easily
- **New service deployment paradigm being developed with OSG & LHC experiments**

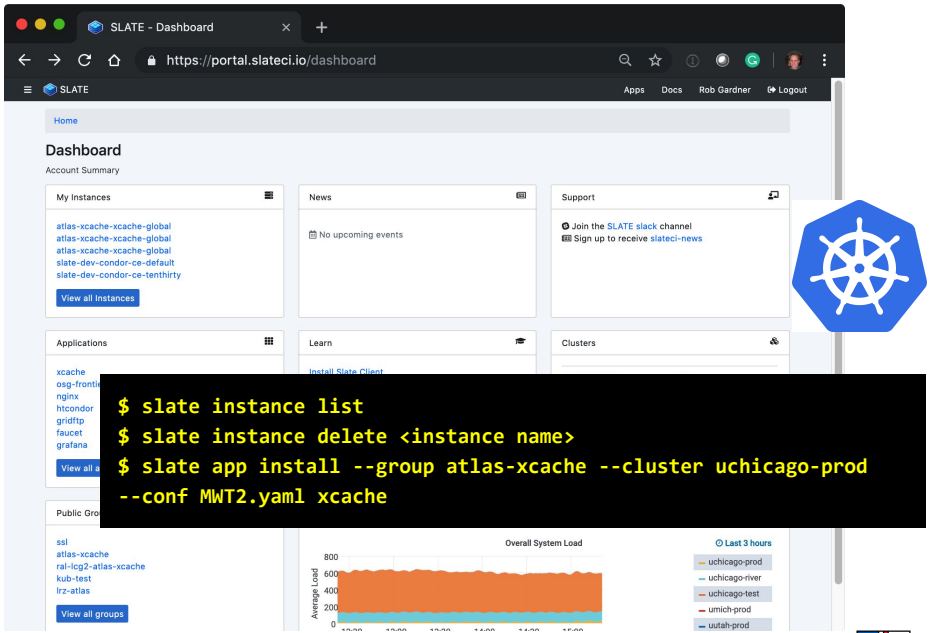
WLCG Worldwide LHC Computing Grid | Open Science Grid

- WLCG SLATE Security working group (with WLCG, trustedci.org, OSG Security & others)

TRUSTED CI THE NSF CYBERSECURITY CENTER OF EXCELLENCE

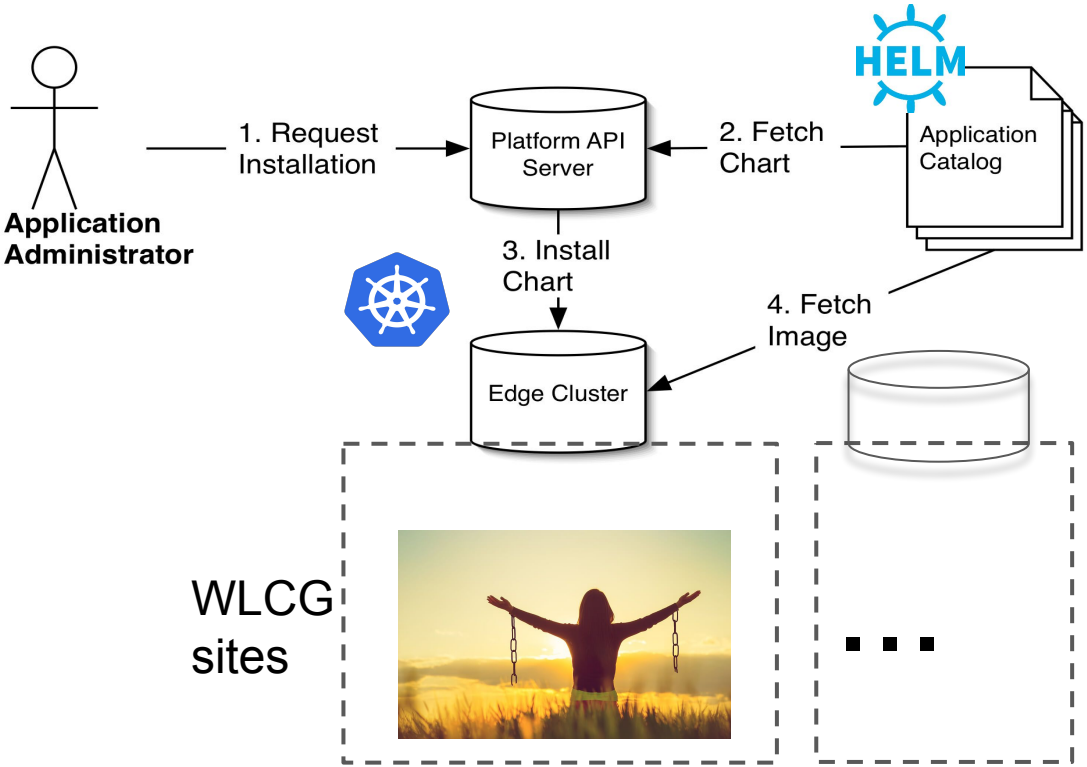
- Edge federation with **Kubernetes**, the industry leading container orchestration platform

- Part of "Facility R&D" needed to inform **next generation Tier2 centers**

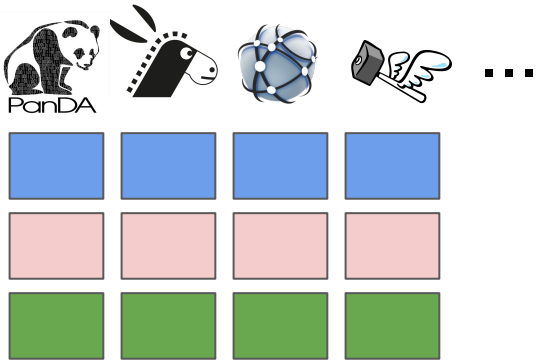


Towards federated NoOps

SLATE CI
slateci.io



ATLAS PLATFORMS



Pacific Research Platform & Chase-CI

NSF OAC-1541349 & NSF CNS-1730158 & NSF OAC-1826967

- A distributed Kubernetes cluster for machine learning and service deployments. Work closely with OSG.
  - More than 400 GPUs across more than 20 universities
    - CMS ML community has accounts for some time.
  - Some FPGAs
    - CMS ML@FPGA development make use of it since 9/2019.
  - TPU in Google Cloud integration
    - CMS ML graduate students explored this, and made relative performance measurements for training on TPU vs GPU to understand TPU utility.
  - Fat network pipes (10-100Gbps) and a network operations team to debug & keep it functional.
  - XRootd cache deployments in US, EU, and Asia at endpoints and in the network backbone.
    - Dominant users include Dune, Minerva, DES and IGWN (LIGO/Virgo/Kagra)
  - ML platform for IRIS-HEP/CoDaS-HEP summer training events (https://codas.slateci.net/)
- Part of the larger NSF-OAC “National Research Platform” vision.

GPU Cloud Bursting in preparation for Exascale

NSF OAC-1941481

- Want to assemble the largest possible cloud GPU burst given the total global GPU capacity for sale by AWS, Azure, and Google.
  - Strong engagement with the cloud providers, Internet2, SDSC, PRP, HTCondor, NSF Cloudbank pilot, OSG, ... and of course IceCube computing operations team.
  - Original goal was 80,000 V100s for 1h …
    - learned since that capacity does not exist today to buy such a 1h burst.
- Do so for IceCube photon simulations
  - Use shower MC as input … i.e. run only the GPU part of the simulation workflow of IceCube
- Larger objective is understanding:
  - HTCondor bursting capability to prepare for “scheduling holes” in Exascale systems of the future to be used for backfill of pre-emptable HTC workflows.
    - Dune, CMS, LIGO/Virgo/Kagra, IceCube, ….
  - Global WAN connectivity between Cloud & on-premise & inter-region within each cloud
- Presenting network measurements at CHEP & GPU burst results at SC19

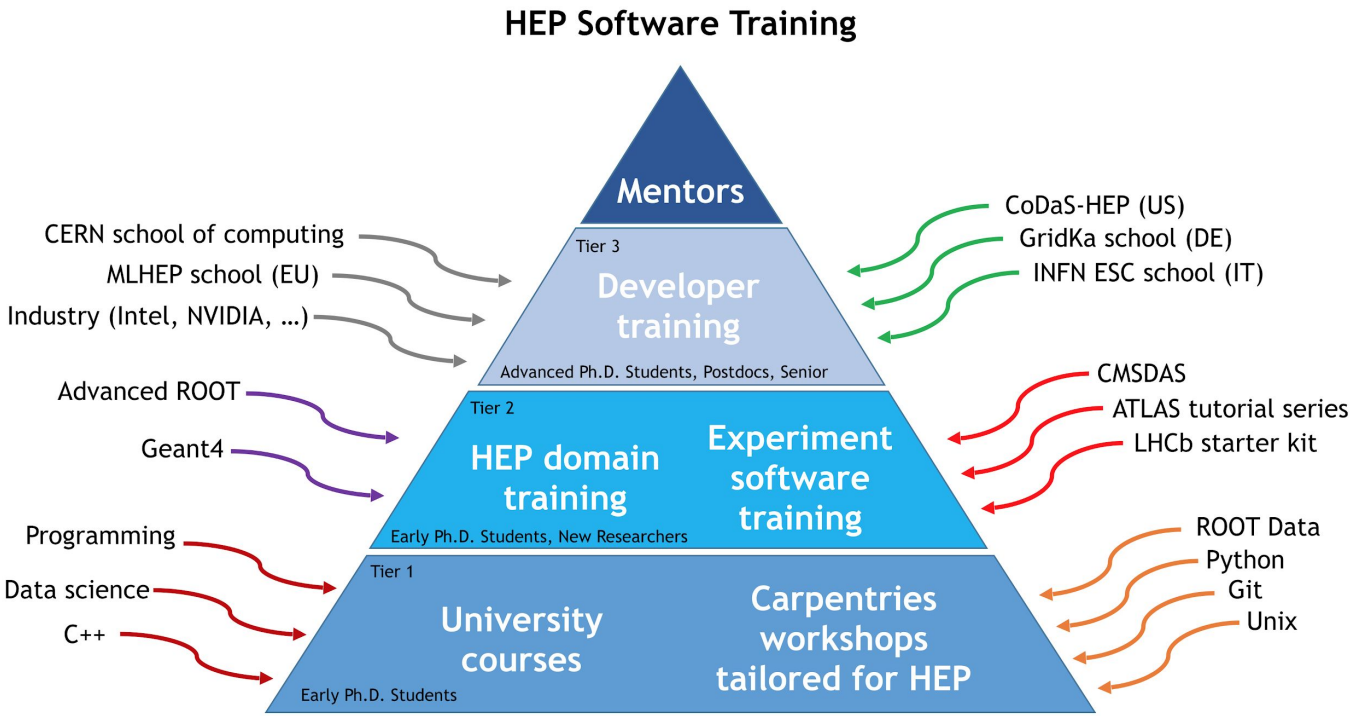
Data Infrastructure for LIGO & IceCube

NSF OAC-1841530

- Goal of this project is to transition LIGO & IceCube to data infrastructure in common with the LHC.
  - Rucio for archival and bulk transfers
  - XRootd & CVMFS as combo for data access towards a global CDN
- Work closely with OSG & PRP
  - DUNE has been the unintended beneficiary of what's been done here.
- Larger Vision:

Better Data Infrastructure via collaboration across the large NSF facilities.

Training and Education - Sustainability/Scalability



This is a general framework for training, but from the NSF we have funds from both IRIS-HEP (OAC-183665) and a separate project FIRST-HEP (NSF OAC-1829707, NSF OAC-1829729, http://first-hep.org) which are working towards implementing this model.

Summary

- SCAILFIN Scalable Cyberinfrastructure for Artificial Intelligence and Likelihood-Free Inference NSF OAC-1841456, 1841471, 1841448
- Advancing Science with Accelerated Machine Learning NSF OAC-1934700, 1934757, 1934360
- PV Finder NSF OAC-1740102 & OAC-1739772
- SAND Service Analysis and Network Diagnosis NSF OAC-1827116
- SLATE Services Layer at the Edge NSF OAC-1724821
- Pacific Research Platform & Chase-CI NSF OAC-1541349 & NSF CNS-1730158 & NSF OAC-1826967
- Data Infrastructure for LIGO & IceCube NSF OAC-1841530
- GPU Cloud Bursting in preparation for Exascale NSF OAC-1941481
- Framework for Integrated Research Software Training (FIRST-HEP) NSF OAC-1829707 & NSF OAC-1829729