

Large-scale HPC deployment of Scalable CyberInfrastructure for Artificial Intelligence and Likelihood Free Inference (SCAILFIN)

Friday, 28 August 2020 11:50 (30 minutes)

High Performance Computing (HPC) facilities provide vast computational power and storage, but generally work on fixed environments designed to address the most common software needs locally, making it challenging for users to bring their own software. To overcome this issue, most HPC facilities have added support for HPC friendly container technologies such as Shifter, Singularity, or CharlieCloud. These different container technologies are all compatible with the more popular Docker containers, however the implementation and use of said containers is different for each HPC friendly container technology. These usage differences can make it difficult for an end user to easily submit and utilize different HPC sites without making adjustments to their workflows and software. This issue is exacerbated when attempting to utilize workflow management software between different sites with differing container technologies.

The NSF-funded Scalable CyberInfrastructure for Artificial Intelligence and Likelihood Free Inference (SCAILFIN) project aims to develop and deploy artificial intelligence (AI) and likelihood-free inference (LFI) techniques and software using scalable cyberinfrastructure (CI) that span multiple sites. Specifically, the project has extended the CERN-based REANA framework, a platform designed to enable analysis reusability, and reproducibility while supporting different workflow engine languages, in order to support submission to different High Performance Computing facilities.

Using the Virtual Clusters for Community Computation (VC3) infrastructure as a starting point, we implemented REANA to work with a number of differing workload managers, including both high performance and high throughput, while simultaneously removing REANA's dependence on Kubernetes support at the workers level.

This work describes the challenges and development efforts involved in extending REANA and the components that were developed in order to enable large scale deployment on HPC resources, as well as the implementation of an abstraction layer that allows the support of different container technologies and different transfer protocols for files and directories between the HPC facility and the REANA cluster edge service from the user's workflow application.

Primary author: Dr HURTADO ANAMPA, Kenyi (University of Notre Dame)

Presenter: Dr HURTADO ANAMPA, Kenyi (University of Notre Dame)

Session Classification: Infrastructure Clouds and Virtualisation Session

Track Classification: Infrastructure Clouds and Virtualisation