Science Federation Emulation Testbed: Demonstration of VFSIE Functionalities

Anees Al-Najjar, Nageswara S. V. Rao, Neena Imam, Seth Hitefield and T. Naughton
Oak Ridge National Laboratory
Oak Ridge, TN 37830, USA
{alnajjaram,raons,imamn,hitefields,naughtont}@ornl.gov

***Abstract*—The Virtual Federated Science Instrument Environment (VFSIE) is a digital twin of a federated infrastructure of multiple sites with science instruments and computing systems that are geographically dispersed. It emulates the federation using containers and virtual hosts (vhosts) connected over local- and wide-area networks for the main purpose of supporting the development and testing of science workflows and software stacks, which will subsequently transition to the physical infrastructure. In this demo, we present VFSIE of a four-site scenario and illustrate the following use cases: (i) execution of federation stack components that expose resources to enable a science user to execute workflows, (ii) execution of remote control commands of beam-line instruments to collect measurements and position the sensor, and (iii) remote execution and access of Jupyter Notebook for tomographic image computations.**

***keywords*: softwarized infrastructures, federated science instruments, containerization, instrument control.**

I. Introduction

Experimental science workflows typically require federations of science instruments and computing systems [4], which are geographically dispersed. These science workflows require complex computations driven by data from the instruments wherein: (i) computations are distributed across the federation's computing platforms, and (ii) instruments are remotely accessed to collect data and steer experiments. These workflows over current federations typically require several months to setup, and subsequently require the continued coordination among multiple organization to operate and maintain.

A software stack, called the FedScI, is proposed by Naughton et al [7] for such a federation to support workflow execution by science users and operations by facility providers. The development and testing of the software stack for the federation and science workflows require the allocations of significant resources and coordination among the sites. It is impractical and ineffective, particularly during early development, to test the stack over production facilities due to the expense and potential disruptions, for example, unintentional out-of-bounds commands to instruments. In response, we proposed the Virtual Federated Science Instrument Environment (VFSIE) [5] which emulates a federation using a combination of containers and virtual hosts (vhosts) connected over an emulated network. In essence, it is a digital twin of the production infrastructure that replicates the federation's software environment by emulating its hosts, networks and instrument targets. VFSIE can be utilized by science users for developing and testing workflow codes and by federation developers and resource providers for testing software stack components used for setting-up and operating the federation.

VFSIE runs within a Virtual Machine (VM)[1] and provides the same software environment as the federation, including science user scripts, instrument control and computational applications packaged as containers and accessed via Jupyter Notebooks. In particular, VFSIE encompasses FedScI for federation resource management and workflow orchestration, Experimental Physics and Industrial Control System (EPICS) for beam-line instrument control demonstration, and imars3d system for computational tools of tomographic imaging accessed via a Jupyter Notebook.

In this demo, we present an emulated federation scenario of four Department Of Energy (DOE) lab sites and demonstrate the following functionalities:

(i) FedScI manager/provider components that expose the federated resources to science users;
(ii) remote access using EPICS system to control a neutron/photon beam-line instrument; and
(iii) remote execution and access to imars3d Jupyter Notebook that generates tomographic reconstructions.

II. VFSIE Design and the use case Scenario

VFSIE emulates a federation of multiple sites connected over a WAN, as shown in Figure 1. Each site is emulated with containers and vhosts representing the federation stack, instrument control, science computation applications, and user codes. The site's components are scattered across multiple computing clusters and instrument stations, and they are connected over different LANs. The emulated sites are connected across the emulated WAN, and the federated network of LANs and the WAN is emulated using ContainerNet [6]. The containers connect to the Docker bridge network (docker0) and also to the federated network to enable access different resources across the federation.

The VFSIE architecture supports the development and testing of science workflows and FedScI modules. In order to validate them, a federation of four DOE laboratory sites, Argonne National Laboratory (ANL), Brookhaven National Laboratory (BNL), National Energy Research Scientific Computing Center (NERSC), and Oak Ridge National Laboratory (ORNL),

[1] VFSIE VM available at https://drive.google.com/drive/folders/1K-7Xcjeh8LfvefqaWeb-r2IwWRC-rV3R?usp=sharing

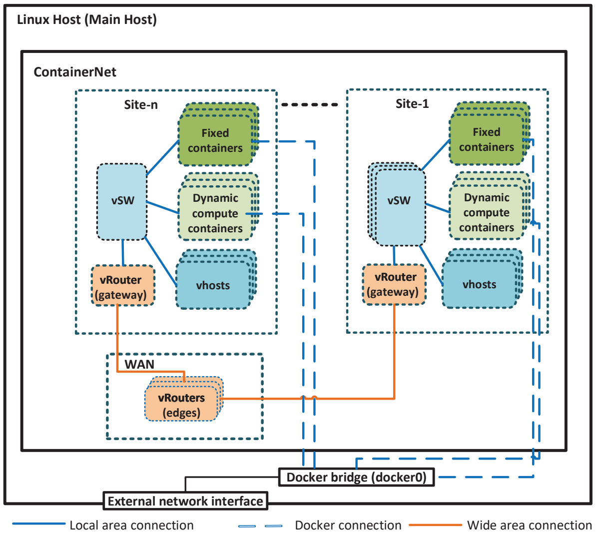


Fig. 1: VFSIE federation infrastructure design.

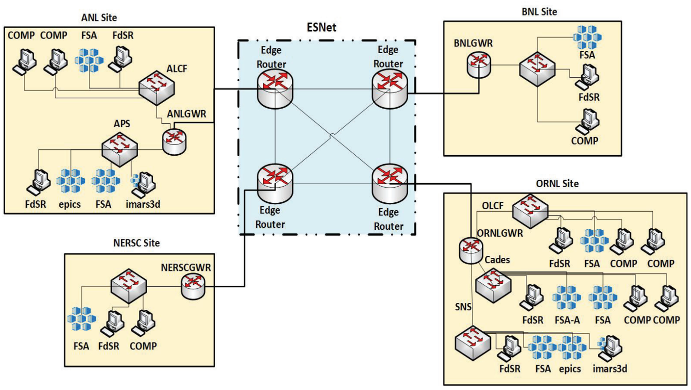


Fig. 2: VFSIE emulation of federation of four DOE sites and their LANs and the WAN connections.

is emulated using VFSIE. These laboratory sites, distributed across the United States, are connected over ESnet [2]. Details of the sites' emulation are shown in Figure 2. The sites provide multiple computing clusters which execute science workflow codes, including Jupyter notebooks packaged as containers, such as imars3d [3]. The beam-line stations at the Advanced Photon Source (APS) at ANL, and the Spallation Neutron Source (SNS) at ORNL provide measurements and also support the positioning of subjects by a science user via EPICS control toolkit. Further details of VFSIE architecture and use case scenarios are available in [5].

III. Demonstration Setup

The demonstration of the emulated federation of the four DOE lab sites illustrates the use of VFSIE for testing and developing FedScI software modules and the science workflow components before deploying over the physical infrastructure. The demonstration consists of two parts: FedScI software stack execution for configuring a federation manager and resource providers, and demonstration of a two-part science workflow use case of controlling a beam-line sensor that generates measurements and accessing the Jupyter Notebook containing codes for reconstructing images using sensor measurements. The FedScI manager runs at the Compute and Data Environment for Science (CADES), and the compute resources are available at CADES and Oak Ridge Leadership Facility (OLCF) at ORNL. For the science workflow, the user is located at BNL and remotely accesses (via SSH) the beam-line sensor and the tomographic reconstruction codes at SNS/ORNL, as shown in Figure 3.

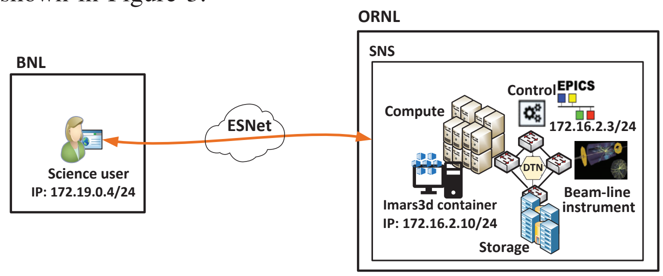


Fig. 3: VFSIE provides a remote access for a science user at BNL to the SNS beam-line instrument at ORNL.

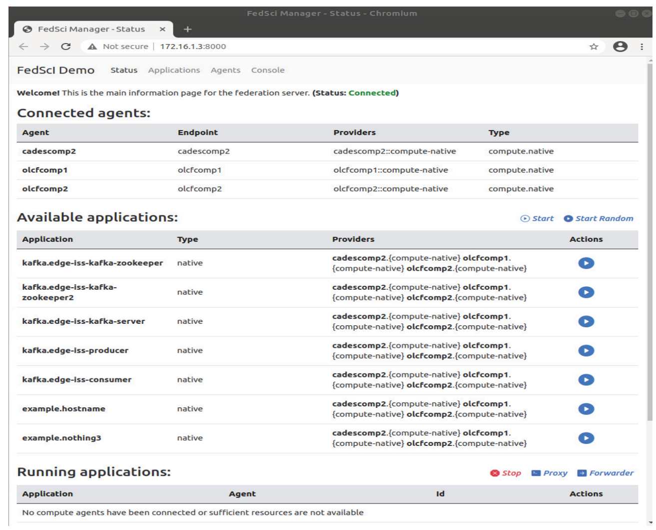


Fig. 4: Access FedScI manager running at CADES cluster at ORNL from BNL site.

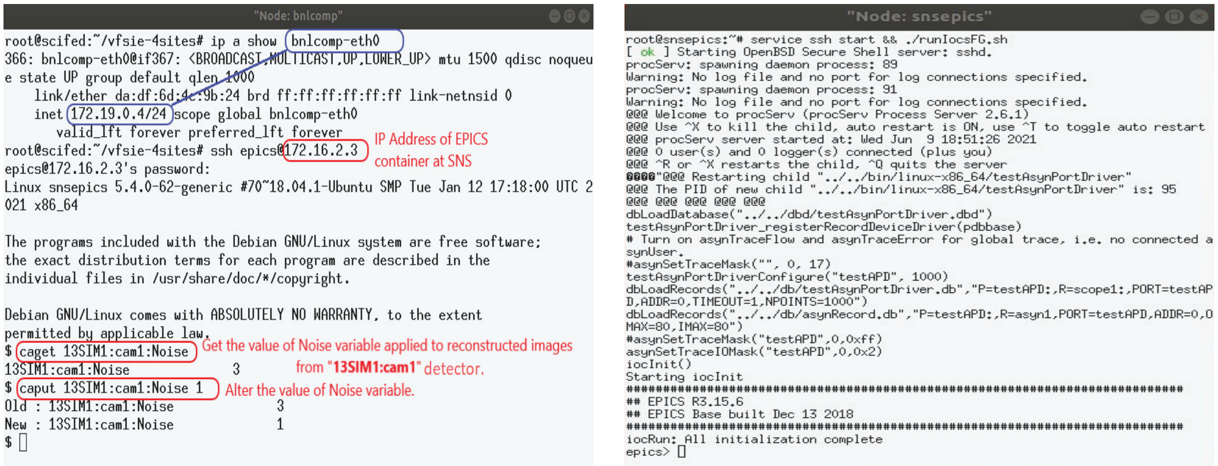
The demo starts with initializing the emulated federation by running VFSIE python script as a command line.

1) Federation manager and resource providers

Once VFSIE is initialized, we run the FedScI manager on the cadescomp1 and the resource providers on cadescomp2, olcfcomp1, and olcfcomp2 hosts. Then, the science user and resource providers access FedScI manager to detect the available resources and execute applications provided by FedScI. This demonstration shows the manager being remotely accessed via Chromium web browser from the BNL site, as illustrated Figure 4.

2) Remote access for instrument control

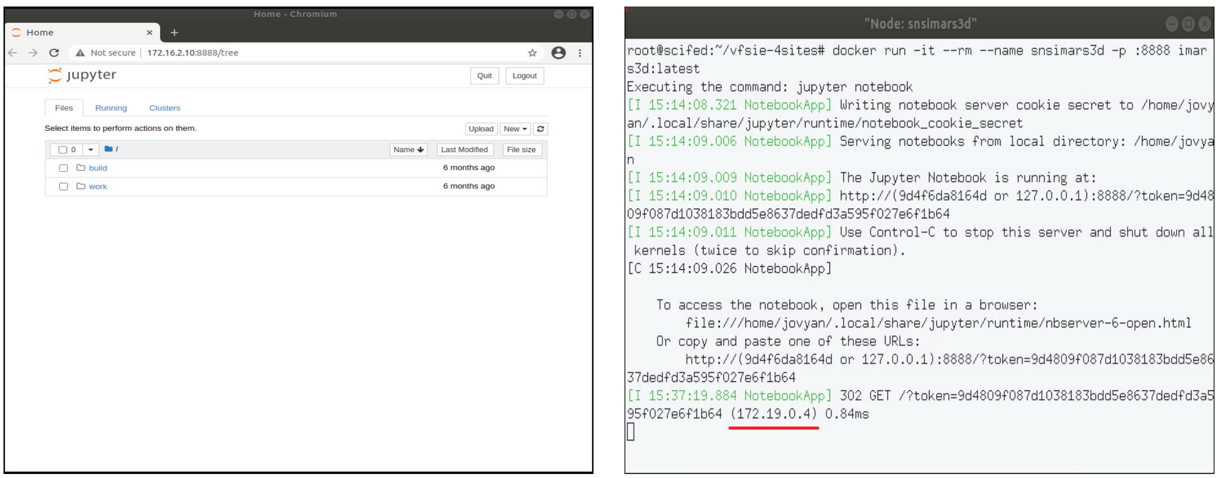
This demonstration shows the science user at BNL remotely accessing an area Detector simulator (Figure 5(a)) running on EPICS [1] at SNS (Figure 5(b)) and configuring the detector variables.



(a) Access EPICS from BNL

(b) EPICS system at SNS for simulating area Detector

Fig. 5: Remote configuration of simulated area Detector running on EPICS container at SNS at ORNL from BNL.



(a) Remote access to imars3d Notebook from BNL

(b) imars3d container at SNS with Get request from BNL

Fig. 6: Access with validation of imars3d Jupyter Notebook at SNS from BNL site.

3) Remote access of data and codes

The last part of this workflow demonstration explains the access to the Jupyter Notebook with imars3d codes available as a container [3] at SNS. Figure 6(a) illustrates access to the notebook via the browser at BNL, and Figure 6(b) shows the running imars3d container at SNS and GET request received from BNL.

IV. Conclusions

The goal of VFSIE is to emulate a federation to support the development and testing of software stack and science workflows in early stages, since running these modules on the production infrastructure is neither always practical nor cost-effective. In this demo, we explained the design of VFSIE, and presented a walk-through of an execution of FedScI stack and a science workflow in a scenario of four DOE lab sites.

V. Acknowledgments

This work is funded in part by SDN-SF and RAMSES projects, Office of Advanced Computing Research, U.S. Department of Energy, and the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC for the U.S. Department of Energy under Contract No. DE-ACO5-00OR22725.

References

[1] EPICS container with simulated IOCs : motor, ADSim and testAsynPortDriver. Available from: https://hub.docker.com/r/lnlssol/epics-sim.
[2] Esnet: Energy sciences network. Available from: https://www.es.net/.
[3] iMars3D: Preprocessing and reconstruction for the Neutron Imaging. Available from: https://github.com/ornlneutronimaging/iMars3D.git.
[4] Software defined networking for extreme-scale science: Data, compute, and instrument facilities. DOE ASCR Workshop, Bethesda, MD.
[5] A. Al-Najjar et al. Virtual framework for development and testing of federation software stack. In 2021 IEEE 46th Conference on Local Computer Networks (LCN), pages 323–326. IEEE, 2021.
[6] M. Peuster et al. Containernet 2.0: A rapid prototyping platform for hybrid service function chains. In 2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft), pages 335–337. IEEE, 2018.
[7] T. Naughton et al. Software framework for federated science instruments. In Driving Scientific and Engineering Discoveries Through the Convergence of HPC, Big Data and AI, pages 189–203. Springer International Publishing, 2020.