Complete End-to-End Network Path Control for Scientific Communities with QoS Capabilities

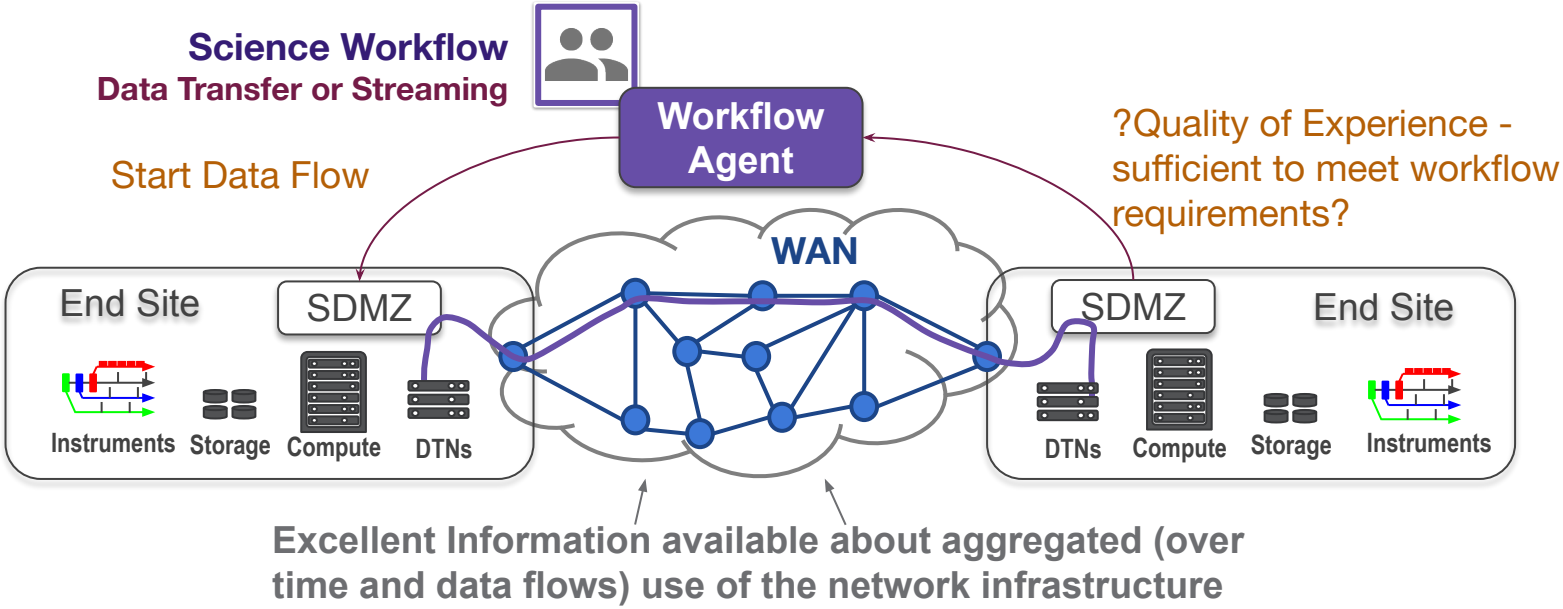
Chin Guok (ESnet)

26th International Conference on Computing in High Energy & Nuclear Physics
May 8 – 12, 2023
Norfolk, Virginia

ESnet

Enable Science Workflow and Network Interaction with Deterministic "Quality of Experience"

- No realtime per flow data available for planning or monitoring
- No "deterministic" network services available
- Start data flow, and hope for the best



Elevate Network to First Class Resource

API driven Automation and Orchestration

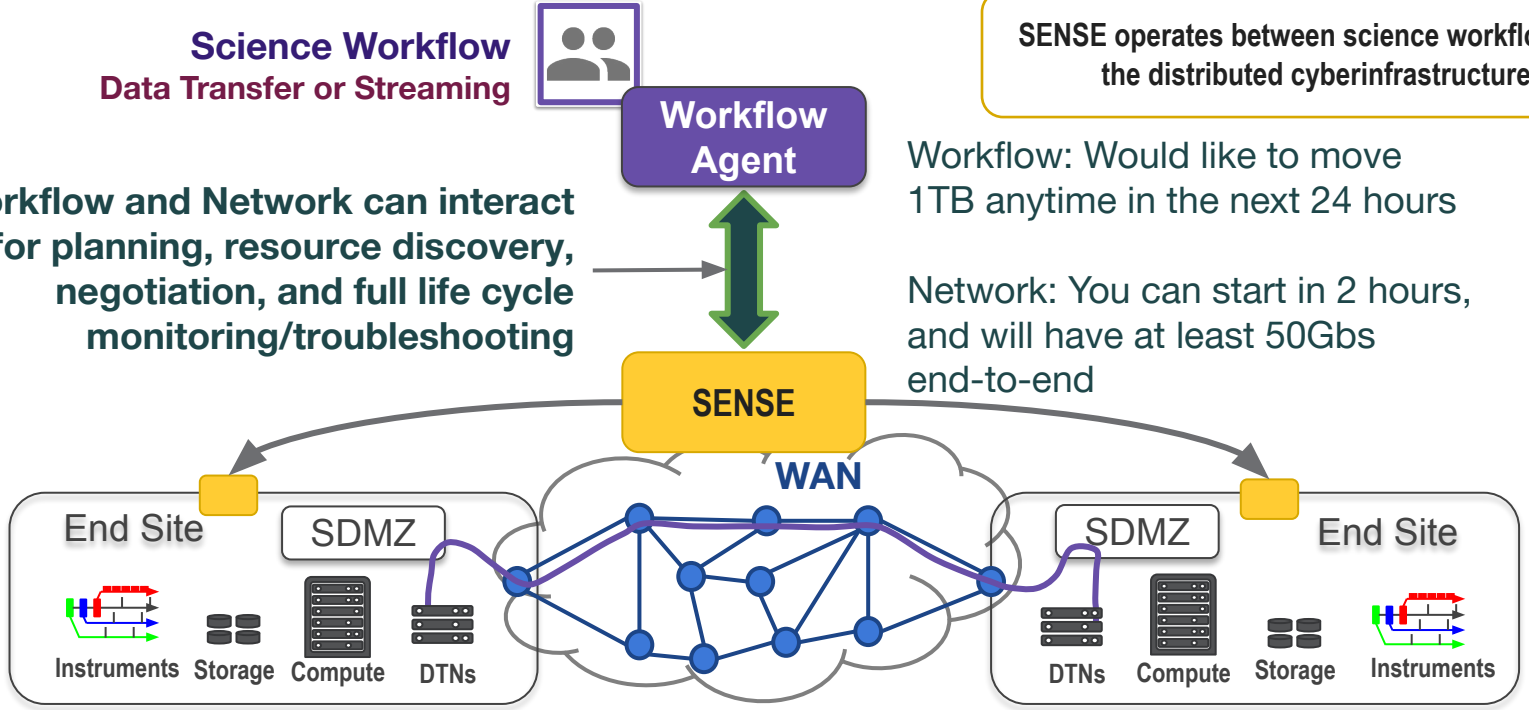
Science Workflow
Data Transfer or Streaming

Workflow and Network can interact for planning, resource discovery, negotiation, and full life cycle monitoring/troubleshooting

SENSE operates between science workflow and the distributed cyberinfrastructure

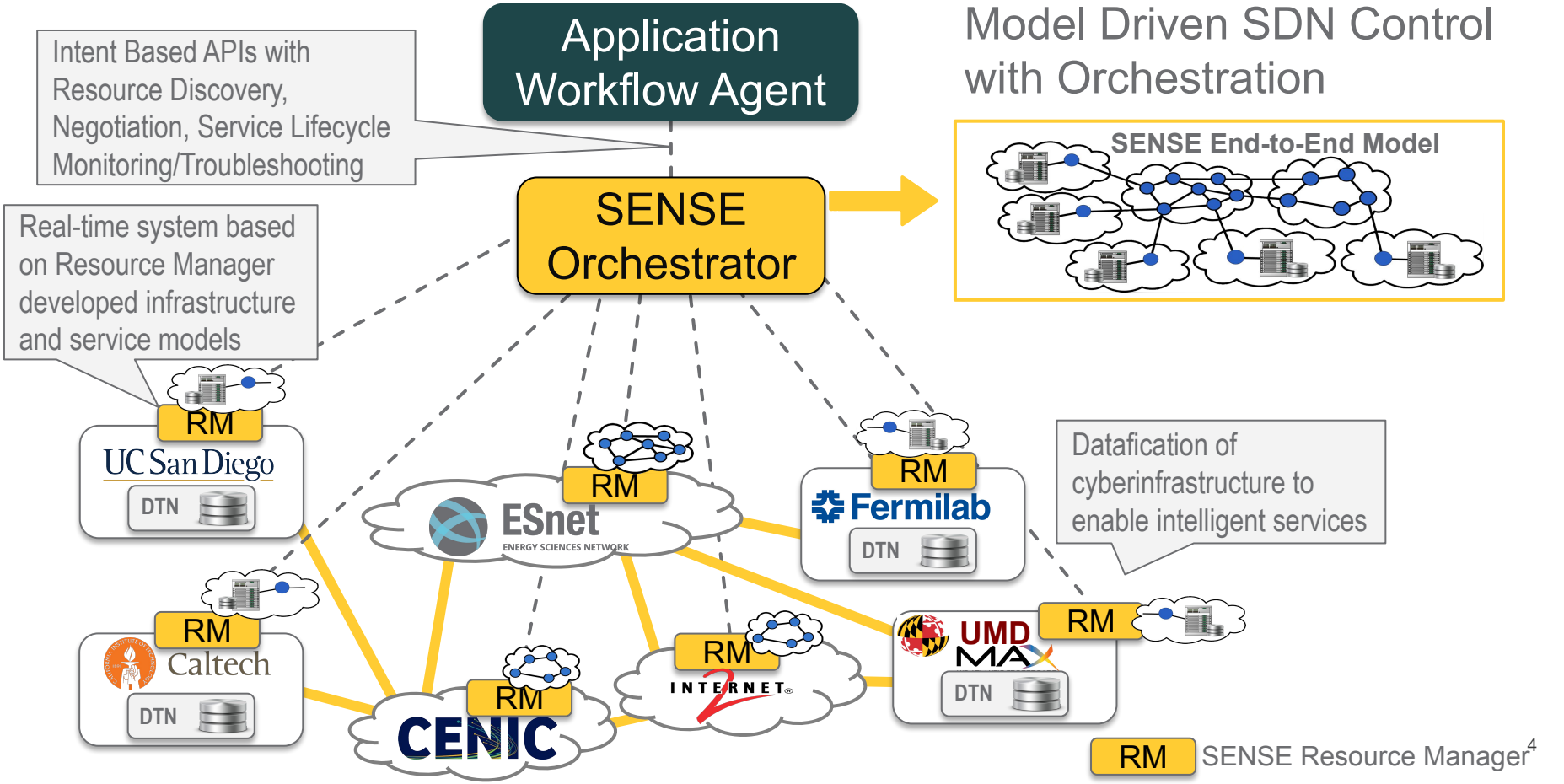
Workflow: Would like to move 1TB anytime in the next 24 hours

Network: You can start in 2 hours, and will have at least 50Gbs end-to-end



- Allows workflows to identify data flows which are higher priority
- Allows the network to traffic engineer to fully utilize all network paths

SENSE Architecture

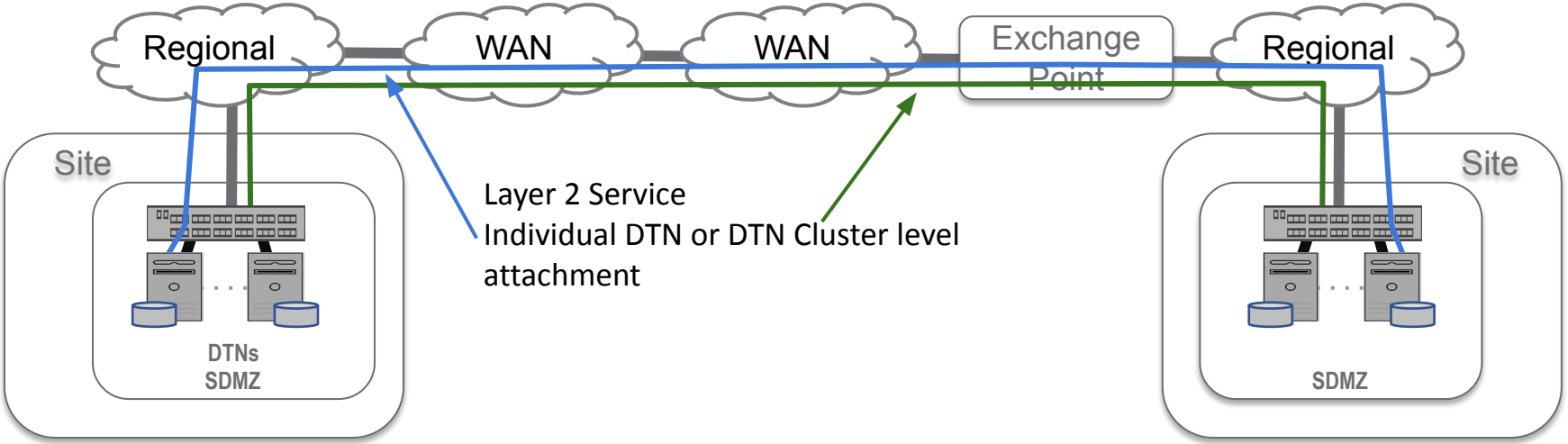


SENSE Solution Approach – Application Interactions

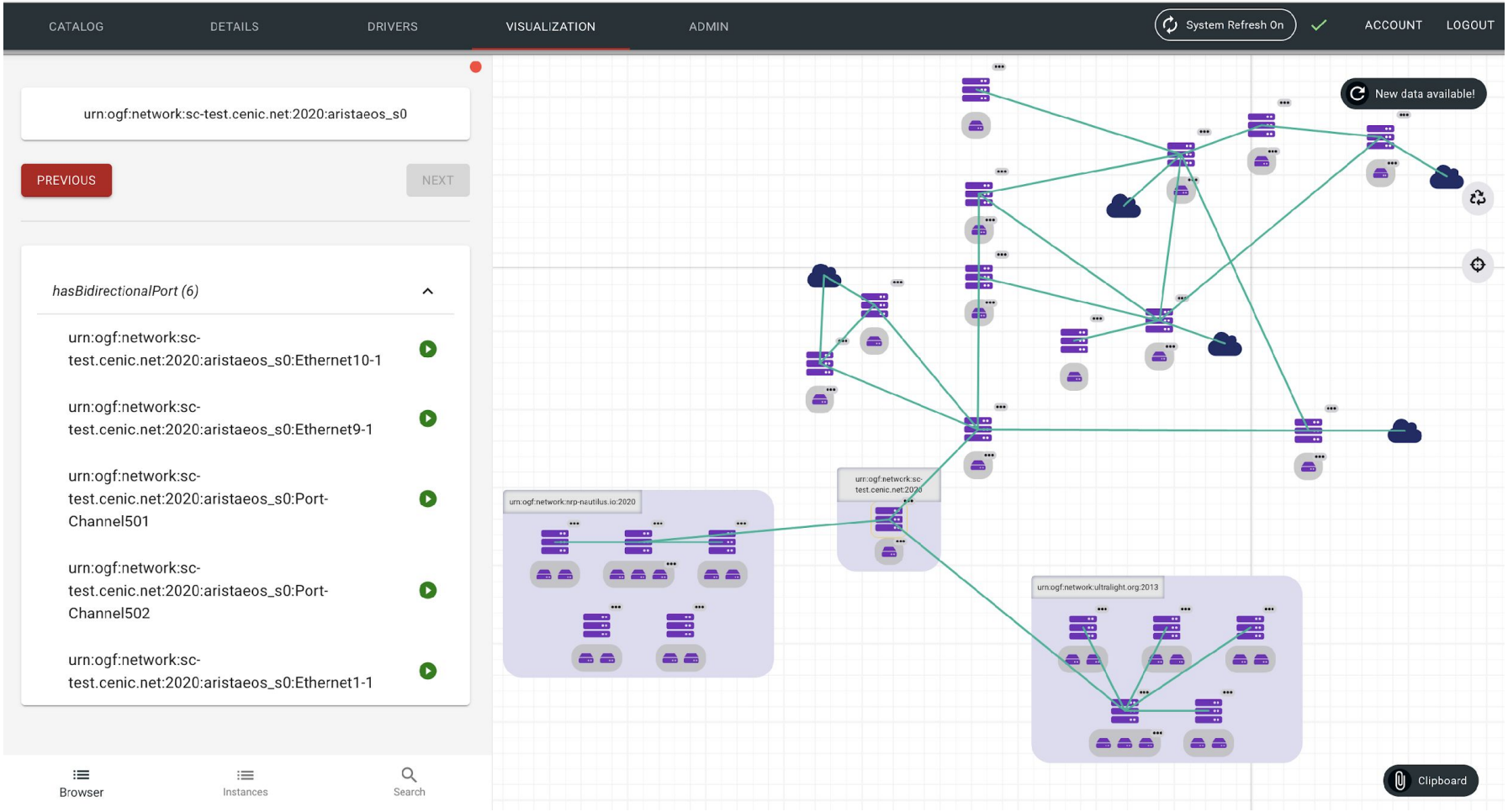
- **Intent Based** – Abstract requests and questions in the context of the application objectives.
- **Interactive** – What is possible? What is recommended? Let's negotiate.
- **Real-time** – Resource availability, provisioning options, service status, troubleshooting.
- **End-to-End** – Multi-domain networks, end sites, and the network stack inside the end systems.
- **Full Service Lifecycle Interactions** – Continuous conversation between application and network for the service duration.

SENSE Services

- Orchestration (of other domain owned systems)
- Multi-Resource (networks, end systems, instruments, clouds)
- Multi-Domain (Sites, Regionals, WANs, Exchange Points)
- Multi-Service (L2 Point-to-Point, L2 MultiPoint, L3VPN, QoS, Traffic engineered paths)
- Intelligent Services, realtime interaction, full-lifecycle monitoring

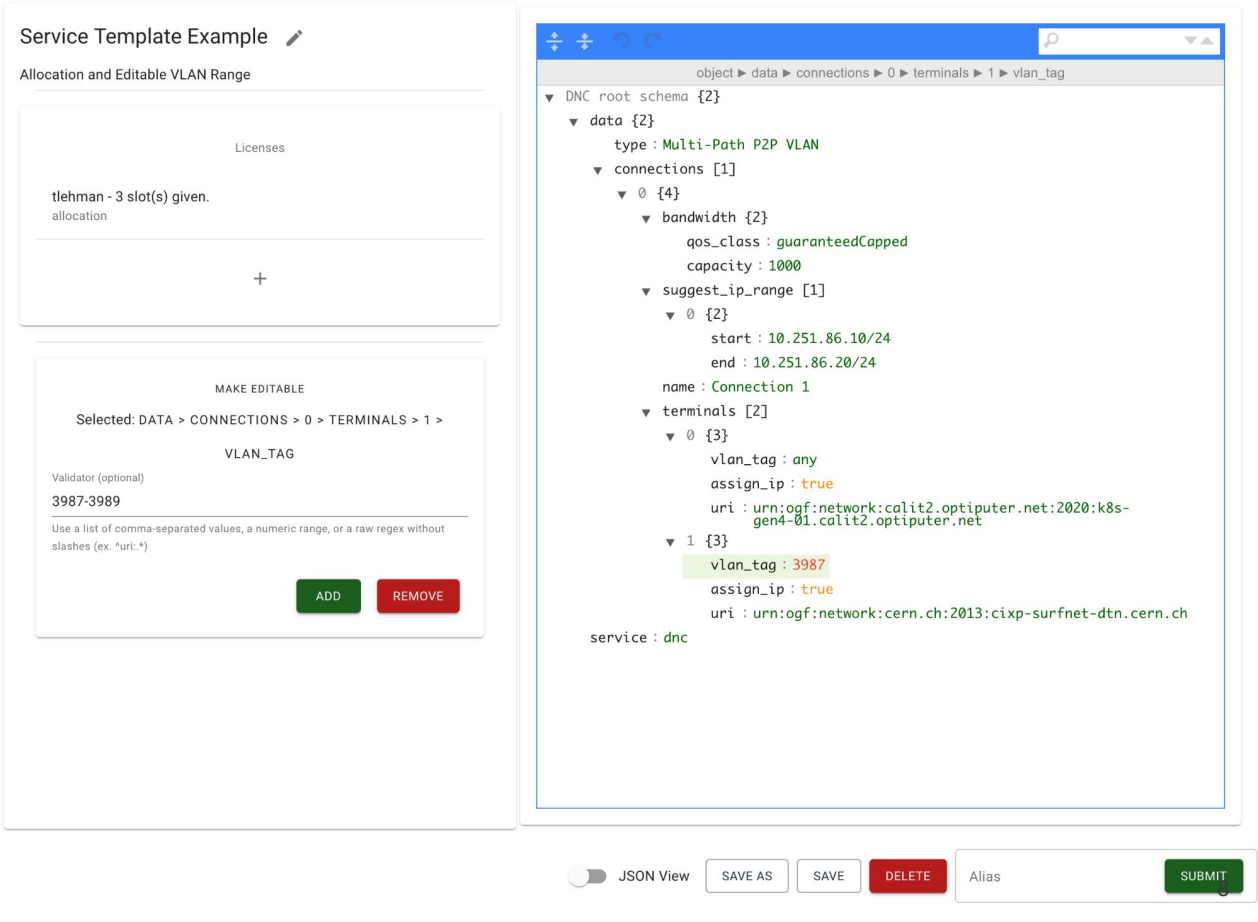


SENSE - Model based Resource Descriptions

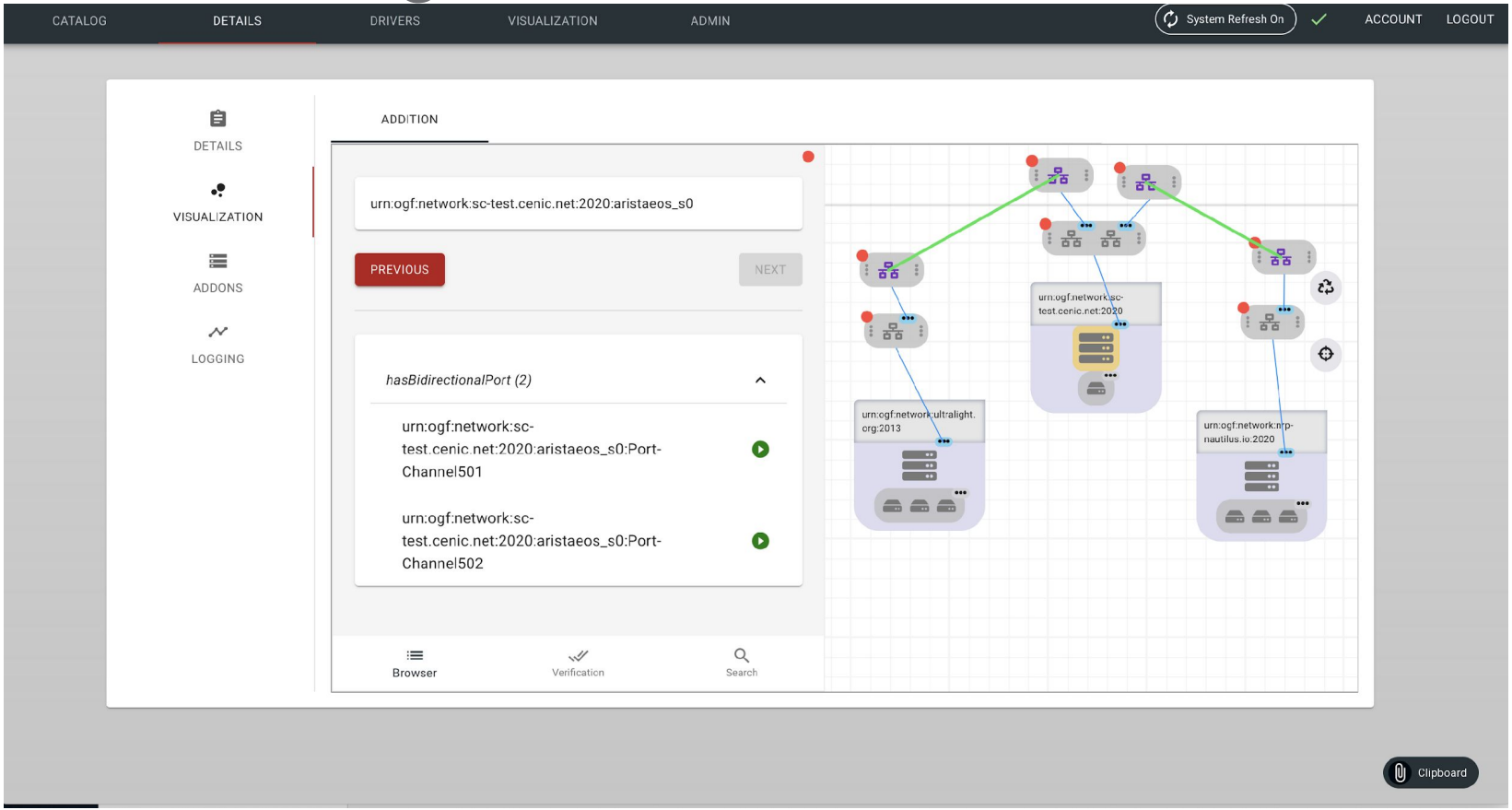


SENSE Orchestrator - Service Template

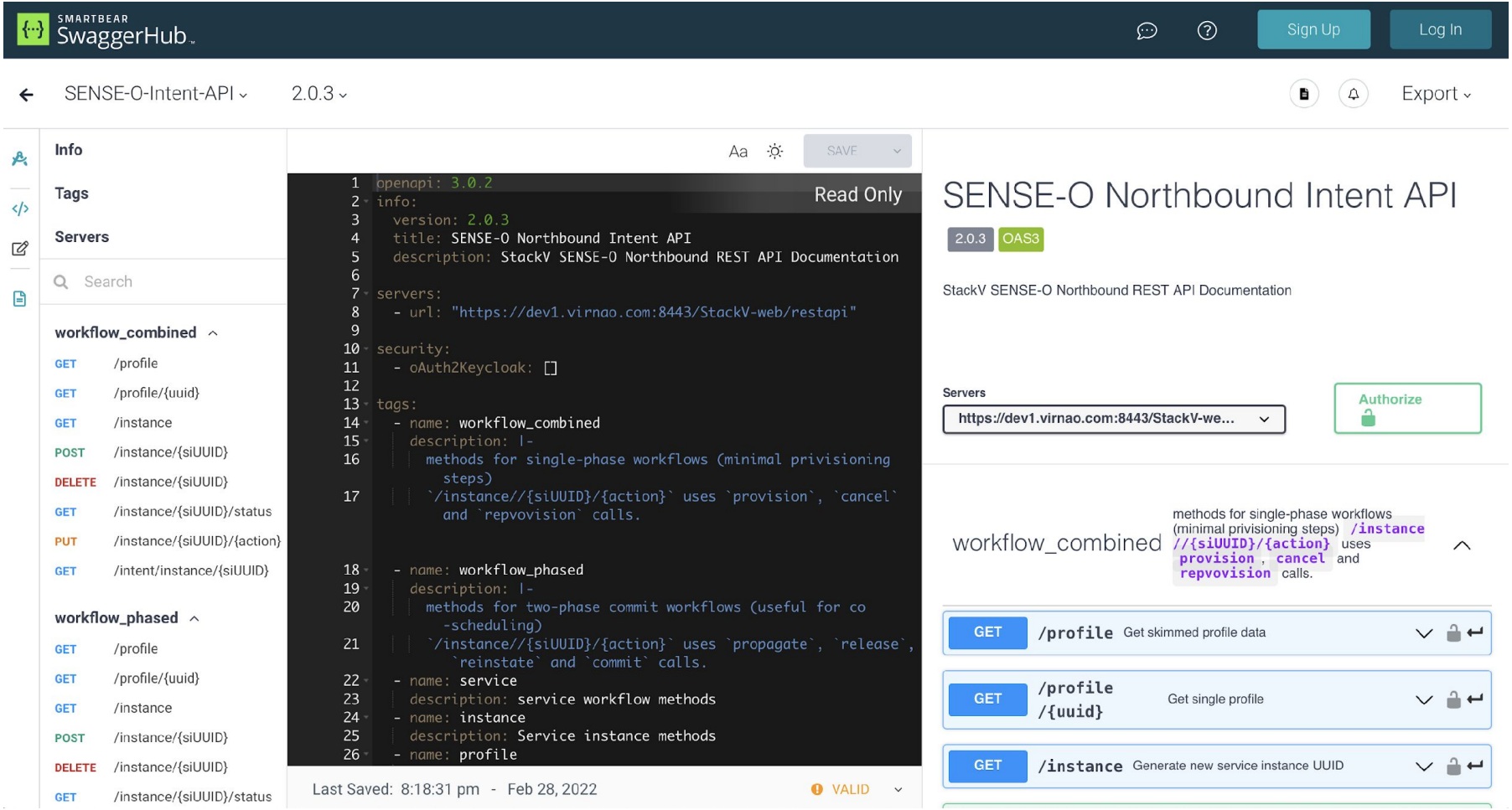
- Read only and optionally with user editable parameters
- Allows user to run with one time "ticket" or multiple time-use allocations



SENSE - Site Layer 3 Flow to WAN Traffic Engineered Path Service

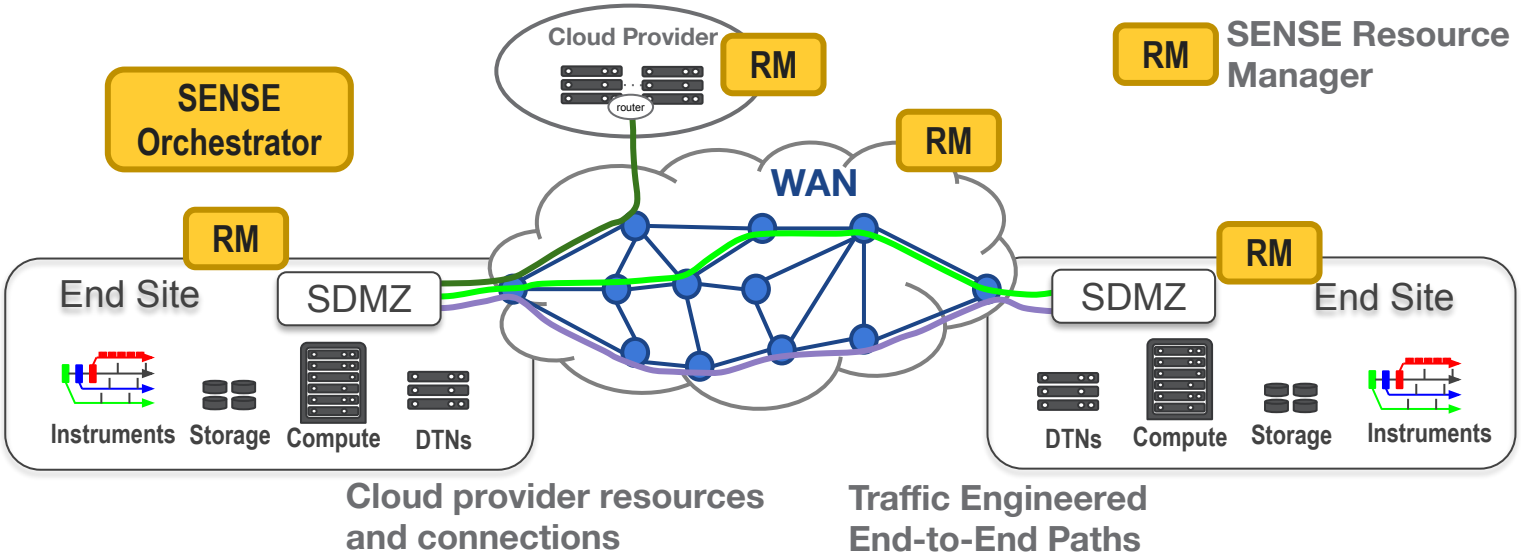


SENSE - Northbound API

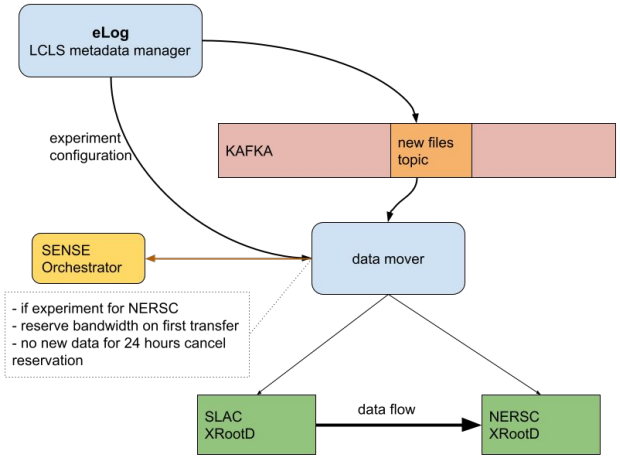


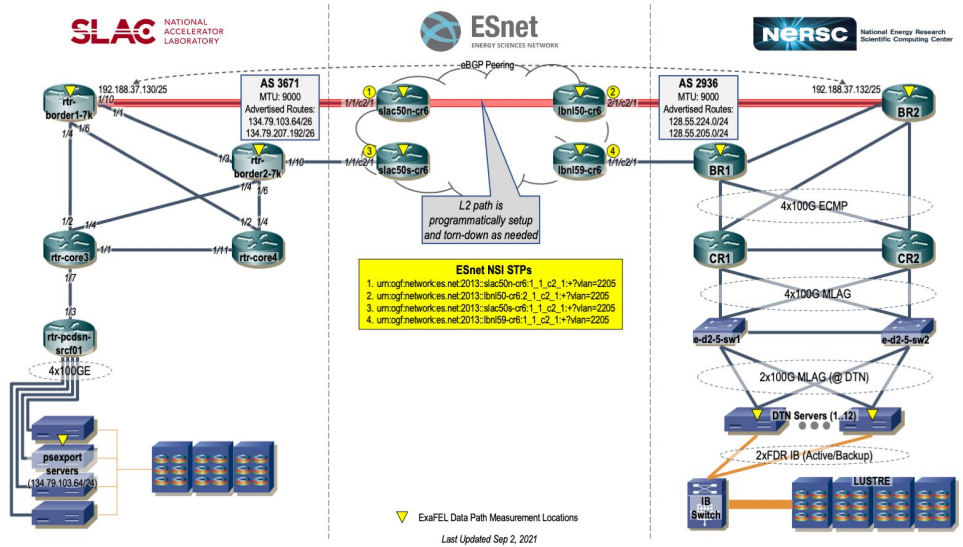
Multi-Resource Orchestration

- Networks, End-Systems, Cloud Resources, Instruments
- No need to manage/orchestrate all of the resources end-to-end, just the ones that matter
  - congestion, performance, or policy reasons



ExaFEL/SuperFacility SENSE Interoperation





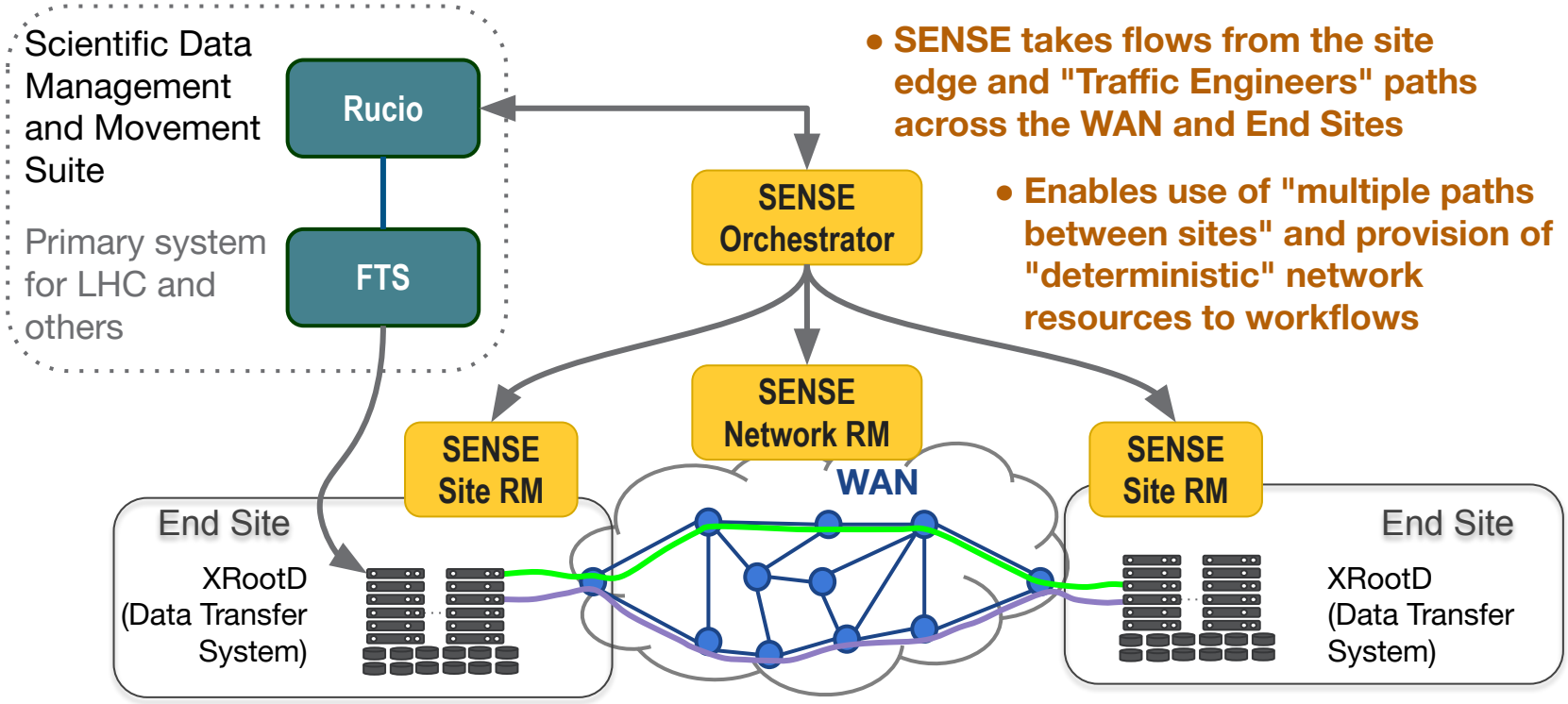
ExaFEL Last-Mile-WG Orchestration Report

October, 2021

Wilko Kroeger[†], Mark Foster[†], Chin Guok[Δ], Damian Hazen[‡], Thomas Lehman[Δ], Ashwin Prabhu Selvarajan[‡], Alexander Sim[º], Xi Yang[Δ]
[‡]NERSC {dhazen, apselvarajan} @ lbl.gov,[Δ]ESnet {chin, tlehman, xiyang} @ es.net,[†]SLAC {fosterm, wilko} @ slac.standford.edu, [º]LBL {asim} @lbl.gov

SENSE and Rucio/FTS/XRootD Interoperation

- **Rucio identifies groups of data flows (IPv6 subnets) which are "high priority"**



- **SENSE takes flows from the site edge and "Traffic Engineers" paths across the WAN and End Sites**
- **Enables use of "multiple paths between sites" and provision of "deterministic" network resources to workflows**

13

SENSE Papers and Info

- Software-Defined Network for End-to-end Networked Science at the Exascale, Elsevier Future Generation Computer Systems, Volume 110, September 2020, Pages 181-201, https://doi.org/10.1016/j.future.2020.04.018
  - Accepted Manuscript: https://arxiv.org/abs/2004.05953
- SENSE Northbound API Program
  - https://app.swaggerhub.com/apis/xi-yang/SENSE-O-Intent-API
- SENSE Website
  - sense.es.net

Future Generation Computer Systems 110 (2020) 181–201

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Software-Defined Network for End-to-end Networked Science at the Exascale

Inder Monga [a], Chin Guok [a], John MacAuley [a], Alex Sim [a], Harvey Newman [b], Justas Balcas [b], Phil DeMar [c], Linda Winkler [d], Tom Lehman [e], Xi Yang [f,*]

[a] Energy Sciences Network, Lawrence Berkeley National Lab, Berkeley, CA, USA
[b] Division of Physics, Mathematics and Astronomy, Caltech, Pasadena, CA, USA
[c] Computing Division, Fermi National Accelerator Laboratory, Batavia, IL, USA
[d] Computing, Environment and Life Science Division, Argonne National Lab, Argonne, IL, USA
[e] Virnao, Arlington, VA, USA
[f] Mid-Atlantic Crossroads, University of Maryland, College Park, MD, USA

ARTICLE INFO

Article history:
Received 1 March 2019
Received in revised form 26 February 2020
Accepted 8 April 2020
Available online 13 April 2020

Keywords:
Intent based networking
End-to-end orchestration
Intelligent network services
Distributed infrastructure
Resource modeling
Software defined networking
Real-time
Interactive

ABSTRACT

Domain science applications and workflow processes are currently forced to view the network as an opaque infrastructure into which they inject data and hope that it emerges at the destination with an acceptable Quality of Experience. There is little ability for applications to interact with the network to exchange information, negotiate performance parameters, discover expected performance metrics, or receive status/troubleshooting information in real time. The work presented here is motivated by a vision for a new smart network and smart application ecosystem that will provide a more deterministic and interactive environment for domain science workflows. The Software-Defined Network for End-to-end Networked Science at Exascale (SENSE) system includes a model-based architecture, implementation, and deployment which enables automated end-to-end network service instantiation across administrative domains. An intent based interface allows applications to express their high-level service requirements, an intelligent orchestrator and resource control systems allow for custom tailoring of scalability and real-time responsiveness based on individual application and infrastructure operator requirements. This allows the science applications to manage the network as a first-class schedulable resource as is the current practice for instruments, compute, and storage systems. Deployment and experiments on production networks and testbeds have validated SENSE functions and performance. Emulation based testing verified the scalability needed to support research and education infrastructures. Key contributions of this work include an architecture definition, reference implementation, and deployment. This provides the basis for further innovation of smart network services to accelerate scientific discovery in the era of big data, cloud computing, machine learning and artificial intelligence.

Published by Elsevier B.V.

1. Introduction

Networked systems are evolving at a rapid pace toward programmatic control, driven in large part by the application of software to networking concepts and technologies, and evolution of the network as a critical subsystem in global scale systems. This is of interest to major science collaborations that incorporate large scale distributed computing and storage subsystems. This software-network innovation cycle is important as it includes a vision and promise for improved automated control, configuration, and operation of such systems, in contrast to the labor-intensive network deployments of today. However, even the most optimistic projections of software adoption and deployment do not put networks on a path that would make them behave as a truly smart or intelligent system from the application or user perspective, nor one capable of interfacing effectively with facilities supporting highly automated data analysis workflows at sites distributed around the world.

Today, domain science applications and workflow processes are forced to view the network as an opaque infrastructure into which they inject data and hope that it emerges at the destination with an acceptable Quality of Experience. There is little ability for

* Corresponding author.
E-mail addresses: imonga@es.net (I. Monga), chin@es.net (C. Guok), macauley@es.net (J. MacAuley), asim@lbl.gov (A. Sim), newman@hep.caltech.edu (H. Newman), jbalcas@caltech.edu (J. Balcas), demar@fnal.gov (P. DeMar), winkler@mcs.anl.gov (L. Winkler), tlehman@virnao.com (T. Lehman), maxyang@umd.edu (X. Yang).

https://doi.org/10.1016/j.future.2020.04.018
0167-739X/Published by Elsevier B.V.

Acknowledgements

Chin Guok, Tom Lehman, Inder Monga, Xi Yang

Harvey Newman, Justas Balcas, Preeti Bhat

Frank Würthwein, Jonathan Guiang, Aashay Arora, Diego Davila, John Graham, Dima Mishin, Thomas Hutton, Igor Sfiligoi

Oliver Gutsche, Phil Demar

15

Thanks