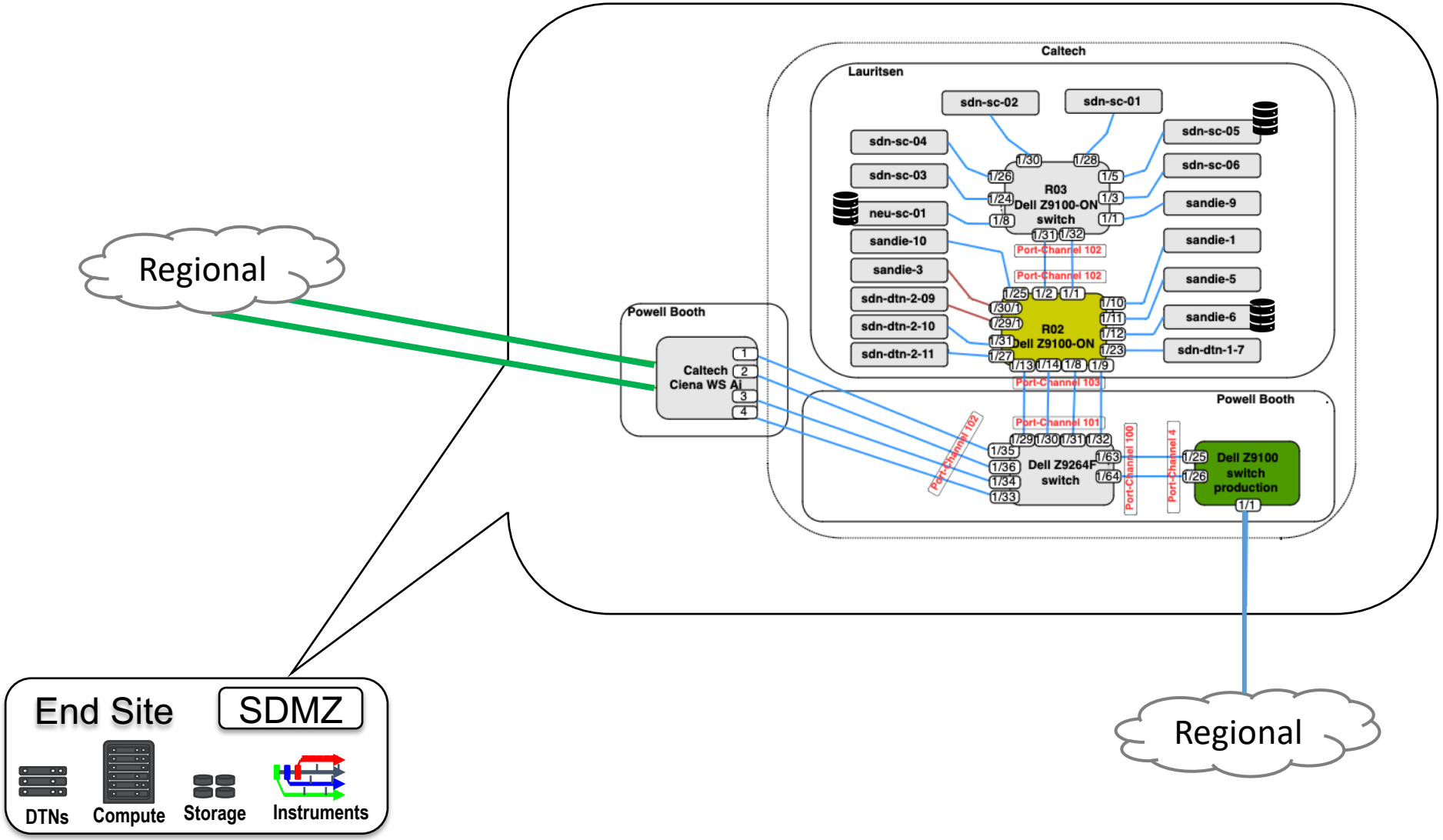
SENSE
SDN for End-to-end Networked Science
At the Exascale

Justas Balcas for SENSE Team
2023, July 12th

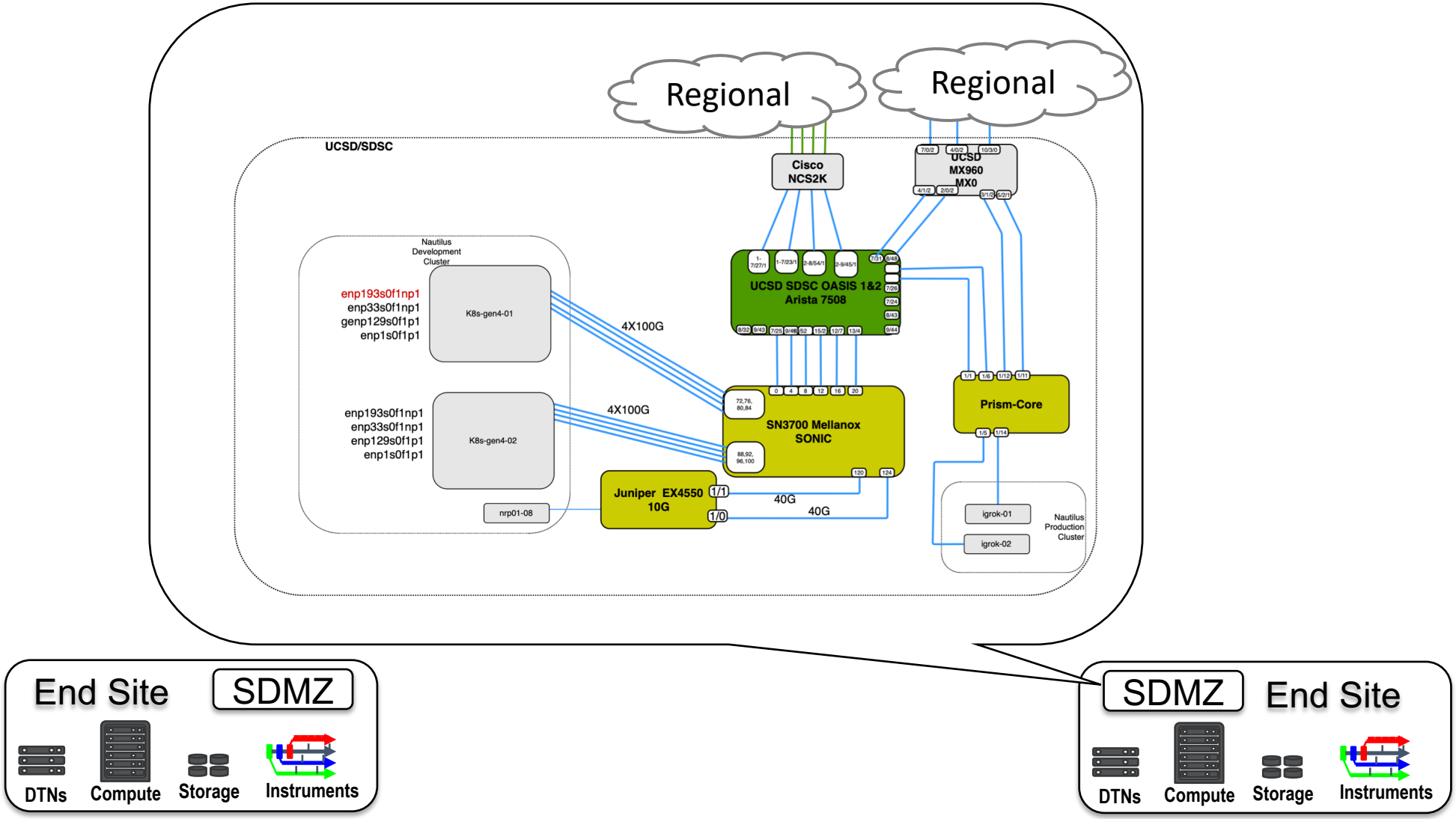
Complexity of control

Caltech



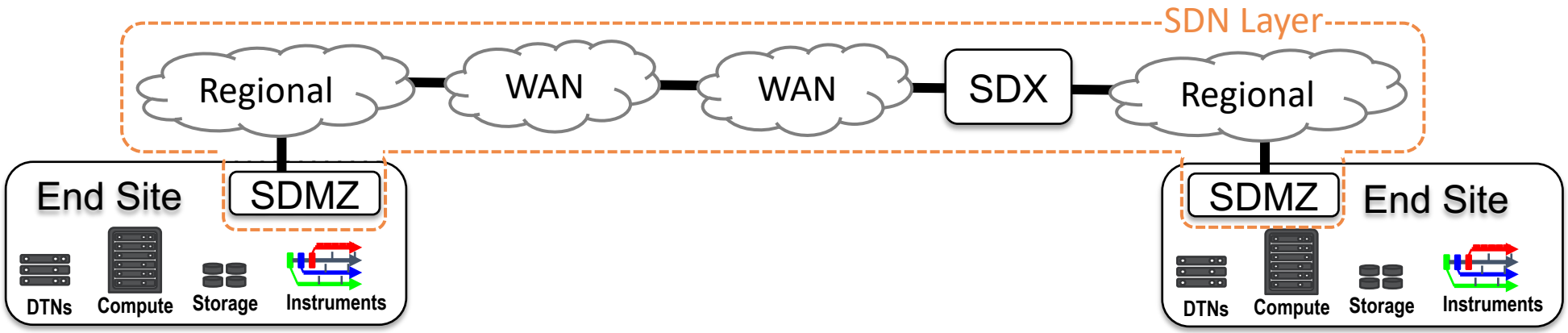
Complexity of control

Caltech



Complexity of control

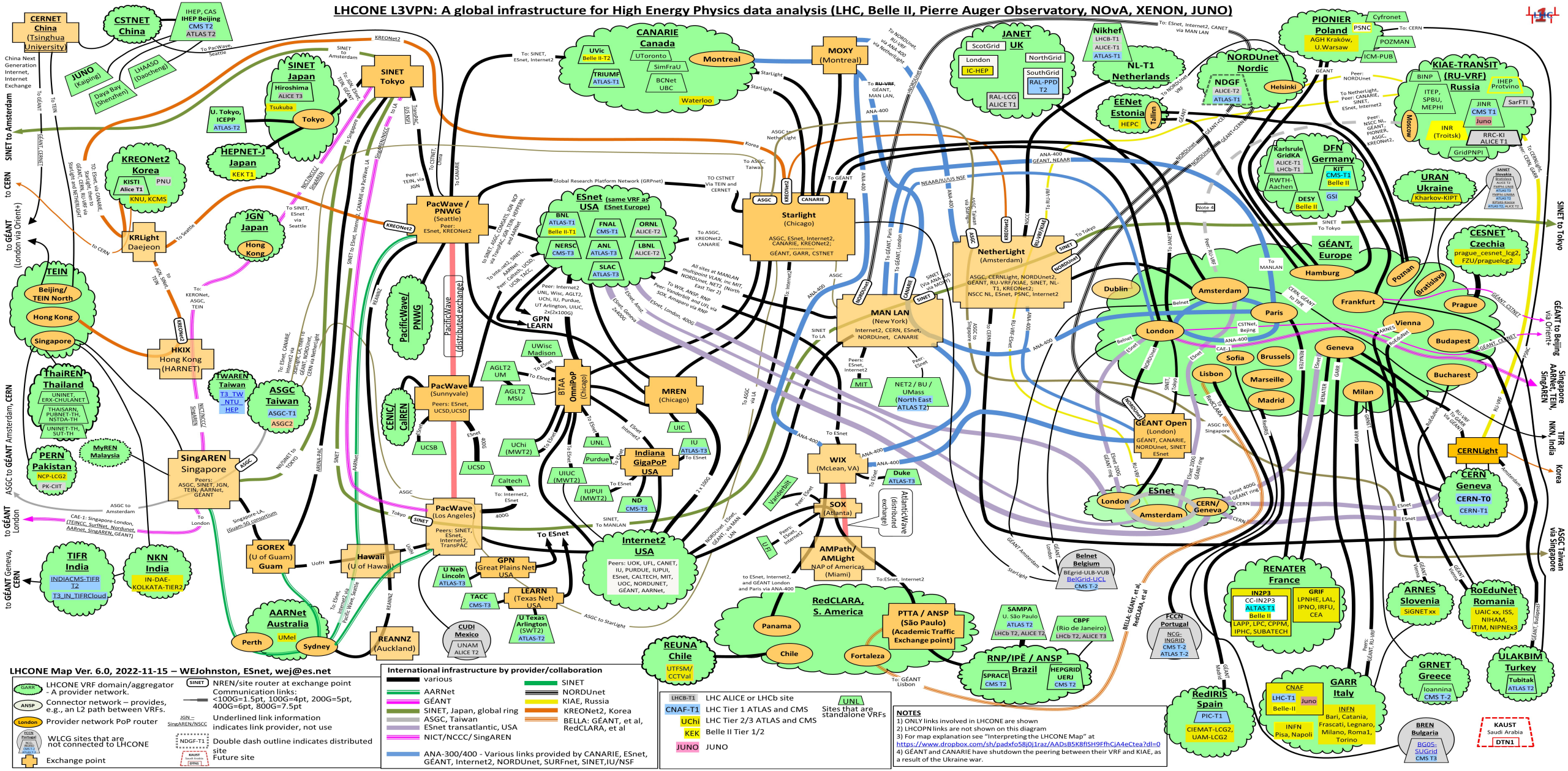
Caltech



LHCONE Network

Map Credit: Bill Johnston

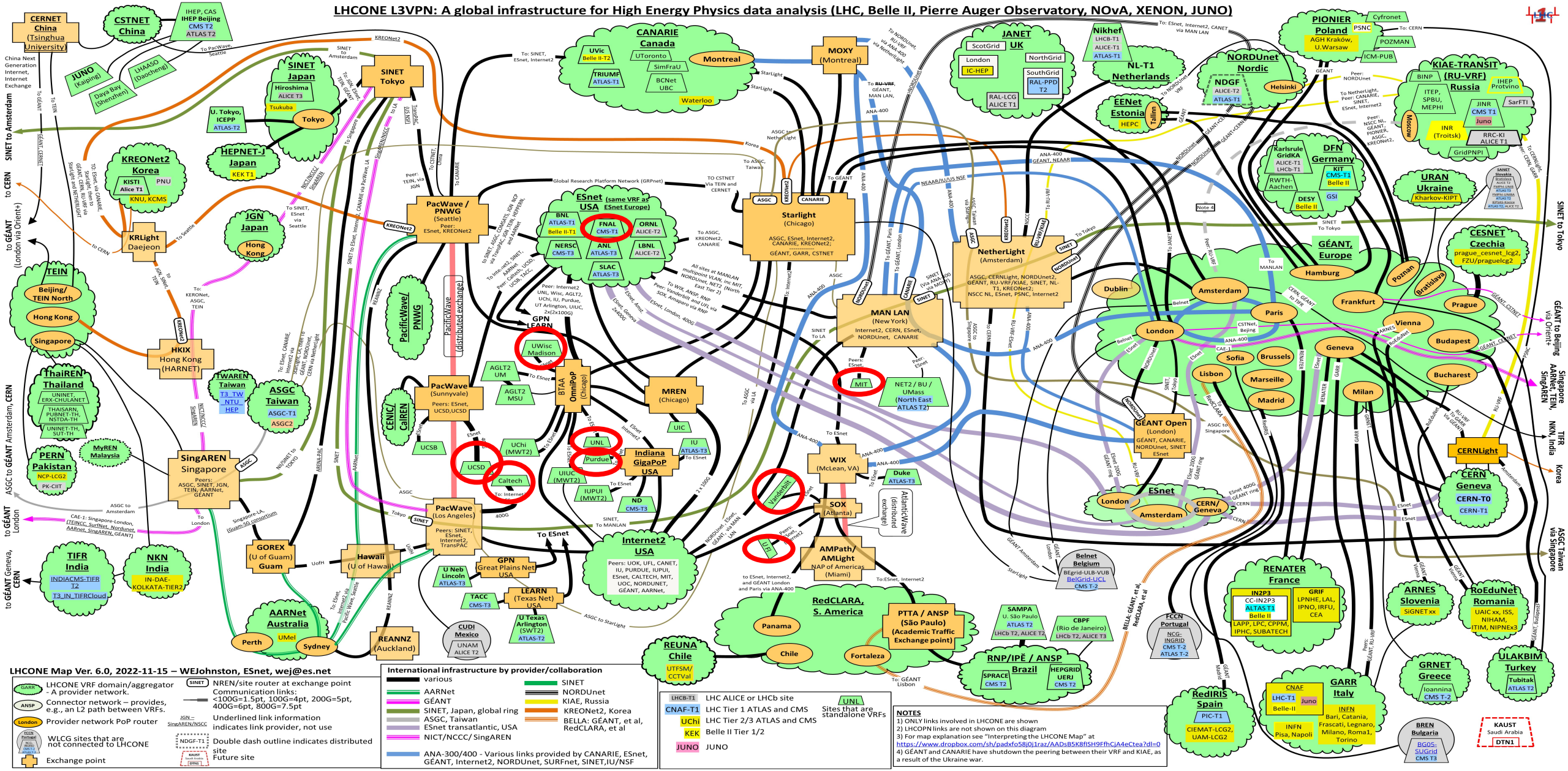
Caltech



LHCONE Network

Map Credit: Bill Johnston

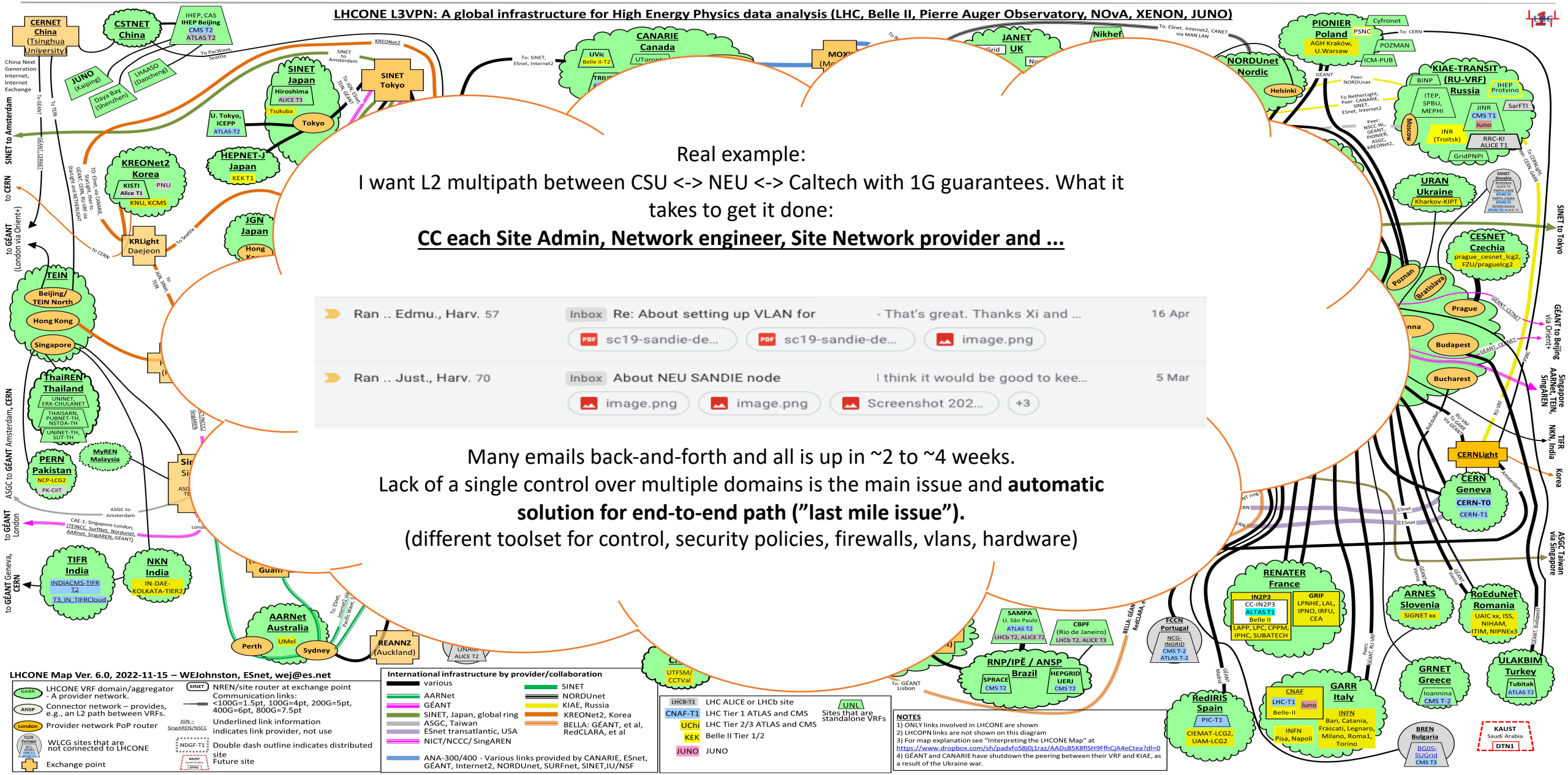
Caltech



LHCONE Network

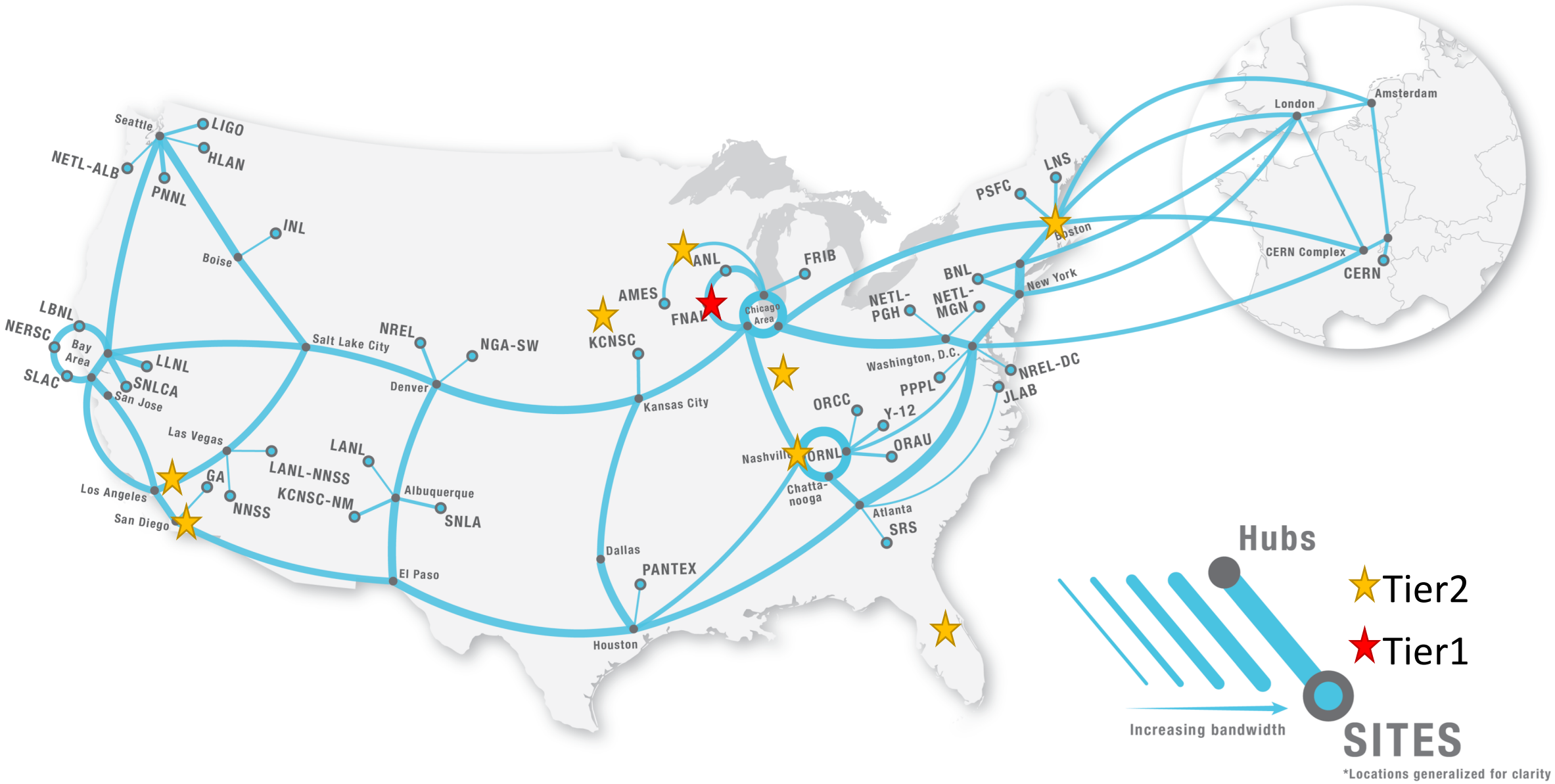
Map Credit:
Bill Johnston

Caltech



ESnet Network Topo + US-CMS Sites

Caltech



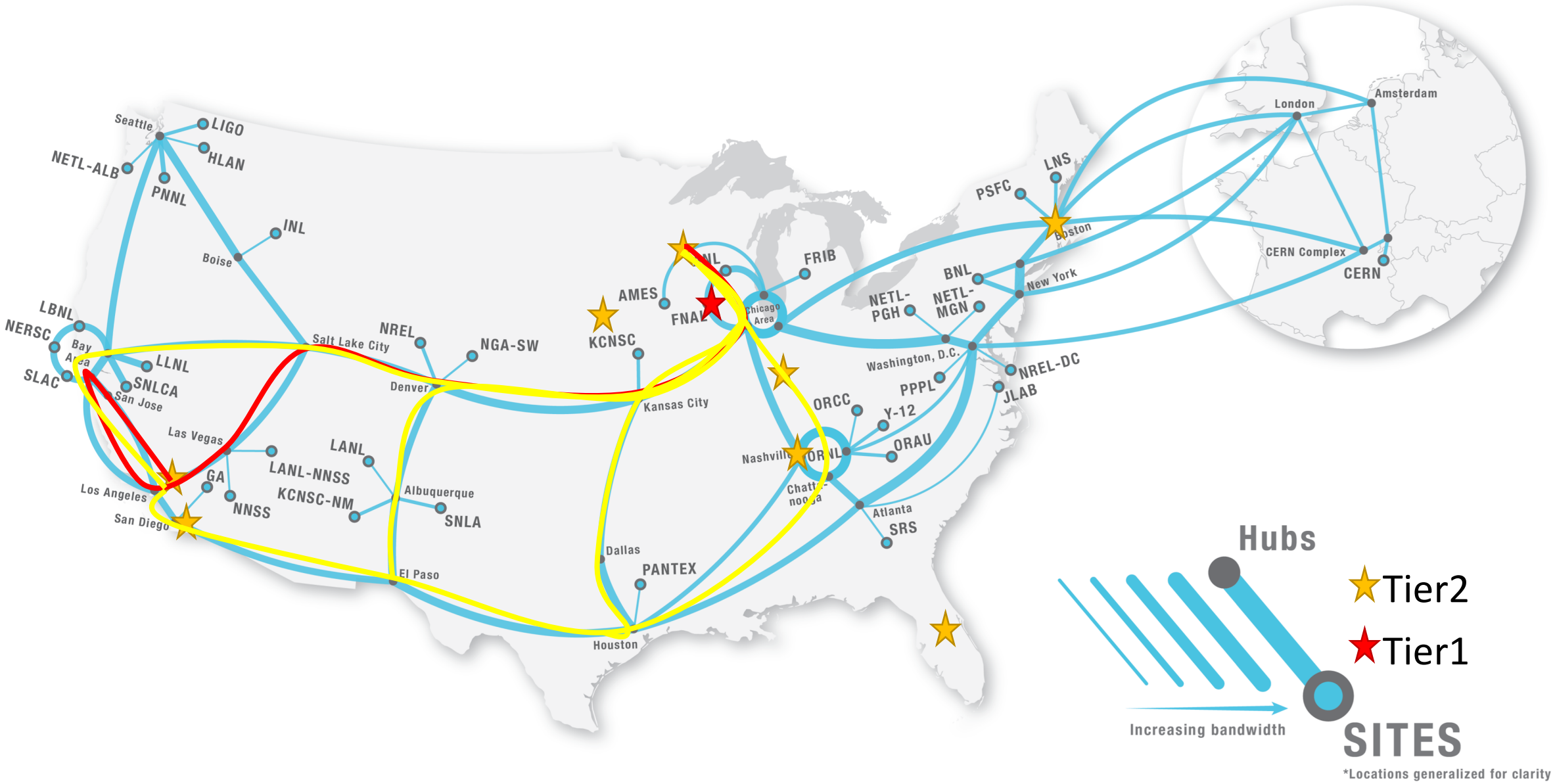
ESnet Network Topo + US-CMS Sites

Caltech

Host	Packets Loss%	Snt	Pings Last	Avg	Best	Wrst	StDev
1. ex01.ultralight.org	0.0%	435	0.3	0.4	0.3	15.1	1.1
2. esnet-lhc1-caltech.es.net	0.0%	435	8.1	8.1	8.0	9.6	0.2
3. losa-cr6--sunn-bb-c.igp.es.net	0.0%	435	7.9	7.9	7.9	9.1	0.1
4. lasv-cr6--losa-bb-c.igp.es.net	0.2%	435	13.0	12.9	12.9	15.0	0.2
5. salt-cr6--lasv-bb-c.igp.es.net	0.0%	434	20.2	20.2	20.1	20.9	0.1
6. denv-cr6--salt-bb-c.igp.es.net	0.0%	434	29.6	29.6	29.5	30.9	0.1
7. kans-cr6--denv-bb-e.igp.es.net	0.0%	434	40.1	40.0	39.9	41.1	0.1
8. chic-cr6--uwmadison_se-203.es.net	0.0%	434	58.3	58.3	58.2	60.2	0.2
9. uwmadison_se-203--chic-cr6.es.net	0.0%	434	75.9	76.7	75.7	117.8	3.8
10. cmsxrootd.hep.wisc.edu	0.0%	434	76.1	76.0	75.9	76.6	0.1

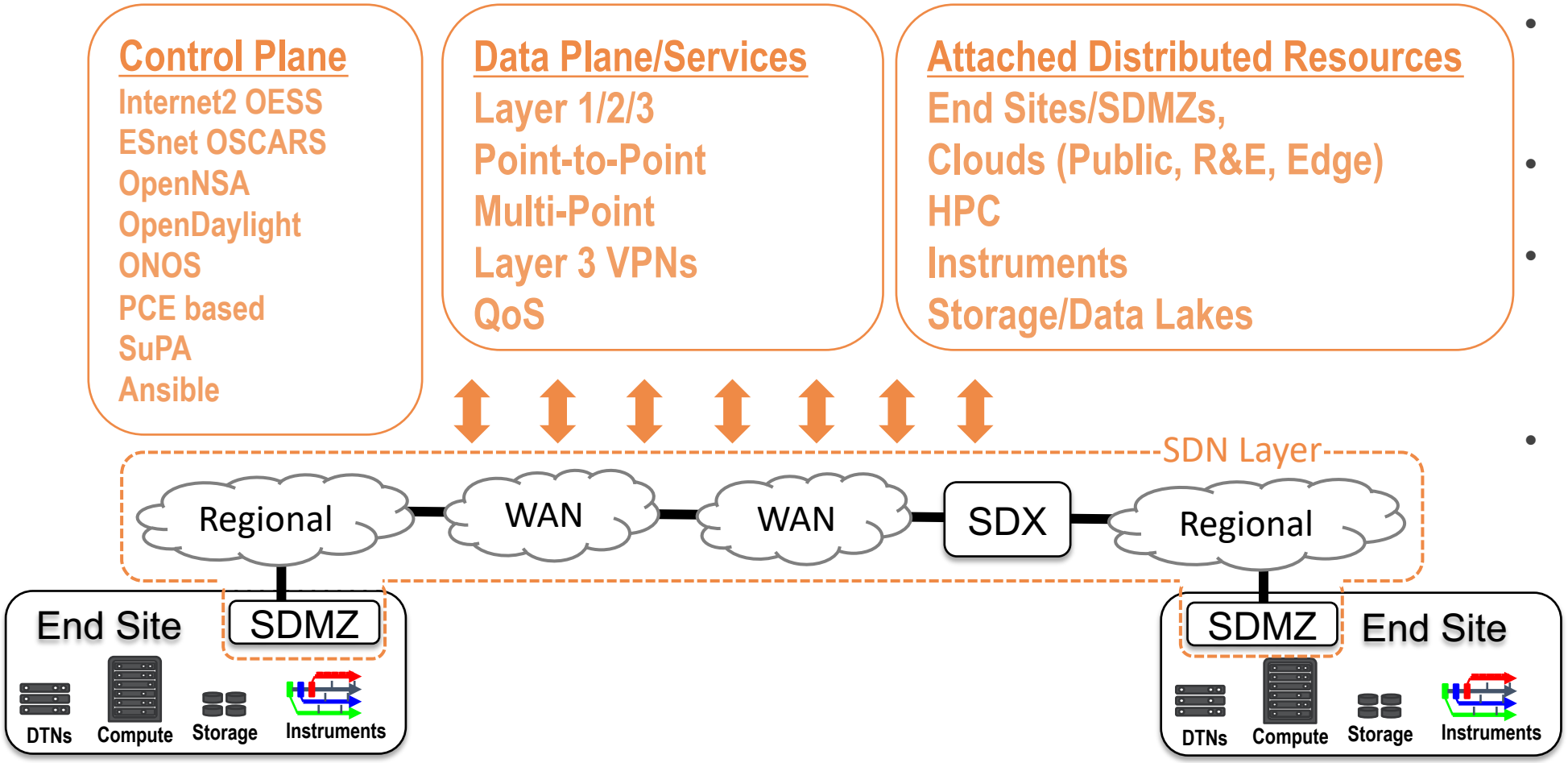
ESnet Network Topo + US-CMS Sites

Caltech



Complexity of control

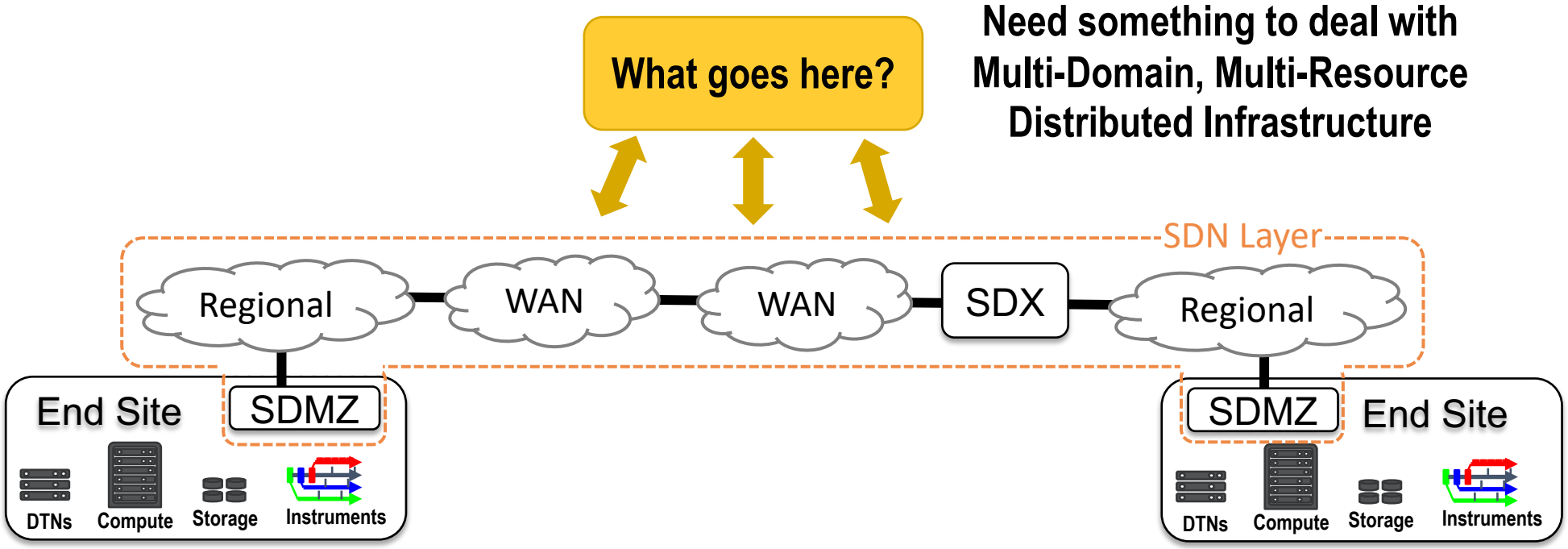
Caltech



- Manual provisioning and infrastructure debugging takes time (a lot)
- No service consistency across domains
- No service visibility or automated troubleshooting across domains
- Lack of real-time information from domains impedes development of intelligent services

Identifying the Gaps

Caltech



SENSE Services

Caltech

- Site Resource Manager (SiteRM)
- Network Resource Manager (N-RM)
- Orchestrator (SENSE-O)

SENSE Service - SiteRM

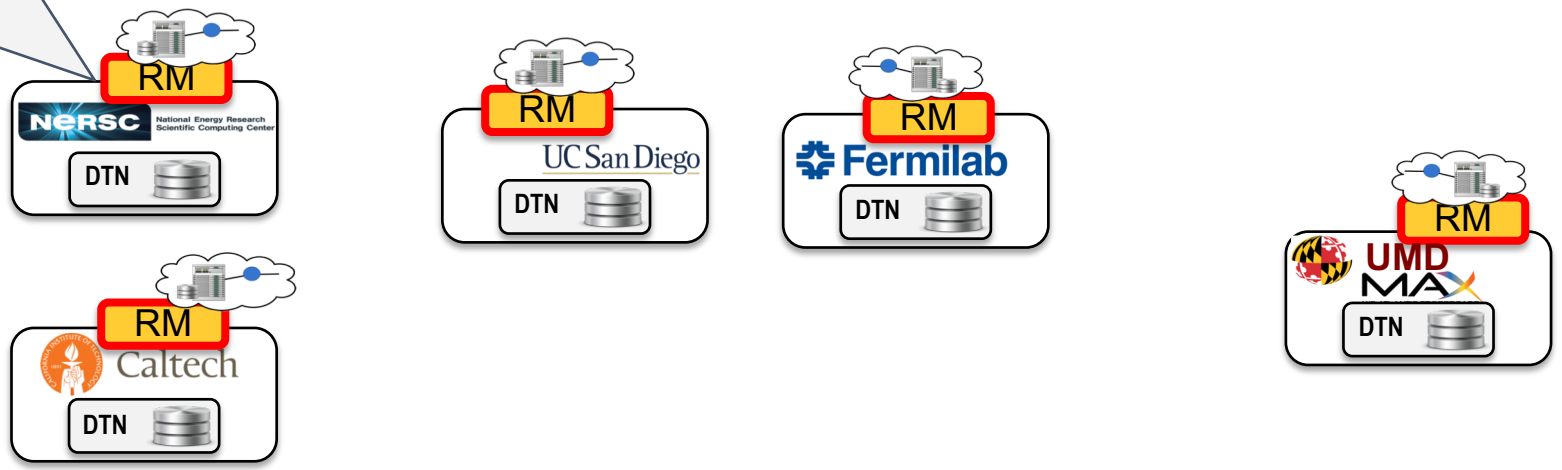
Caltech

SiteRM Functions/Roles:

- Responsible for a specific set of Site, Network and Node Resources
- Generate real-time MRML Model
- Evaluate and respond to SENSE Orchestrator information and service requests (including negotiation)
- Provision Site resources in support of SENSE services (includes networking stack of end systems)
- QoS provided via Traffic Control (FireQoS/tc)
- Automatic dataflow initiation for path verification
- Real time monitoring of Network and Node Resources (Prometheus format)
- Debug actions (Ping, Traceroute, Throughput Test)

Deployment Requirements:

1. Docker/Podman on each controlled end-host
   - NET_ADMIN privileges
2. Run siterm-agent on each host in container
3. Access to network devices (SiteRM uses ansible)
4. VLAN and IPv4/IPv6 private range allocation
5. SNMP Data from all network devices
6. Node Exporter on all controlled end-hosts



SENSE Service – Network RM

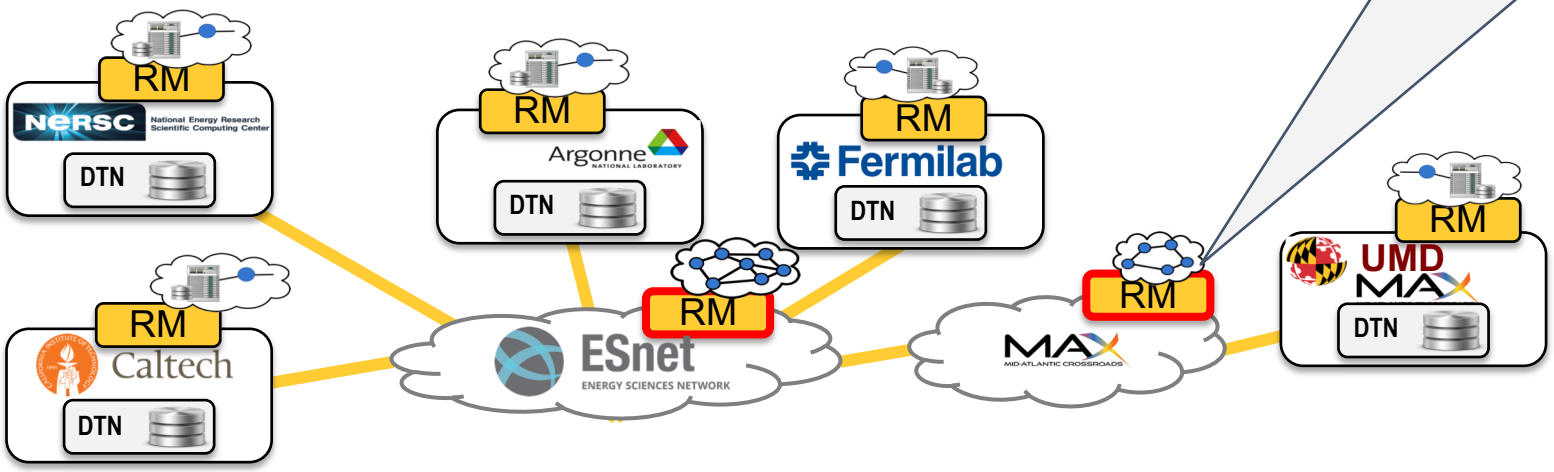
Caltech

Deployment Requirements:

1. Docker/Podman/VM
2. Run any supported Network RM Stack (OSCARS, OpenNSA, SuPA, etc…)
3. Access to network devices
4. SNMP Data from all network devices (or other way export port statistics, like Stardust)

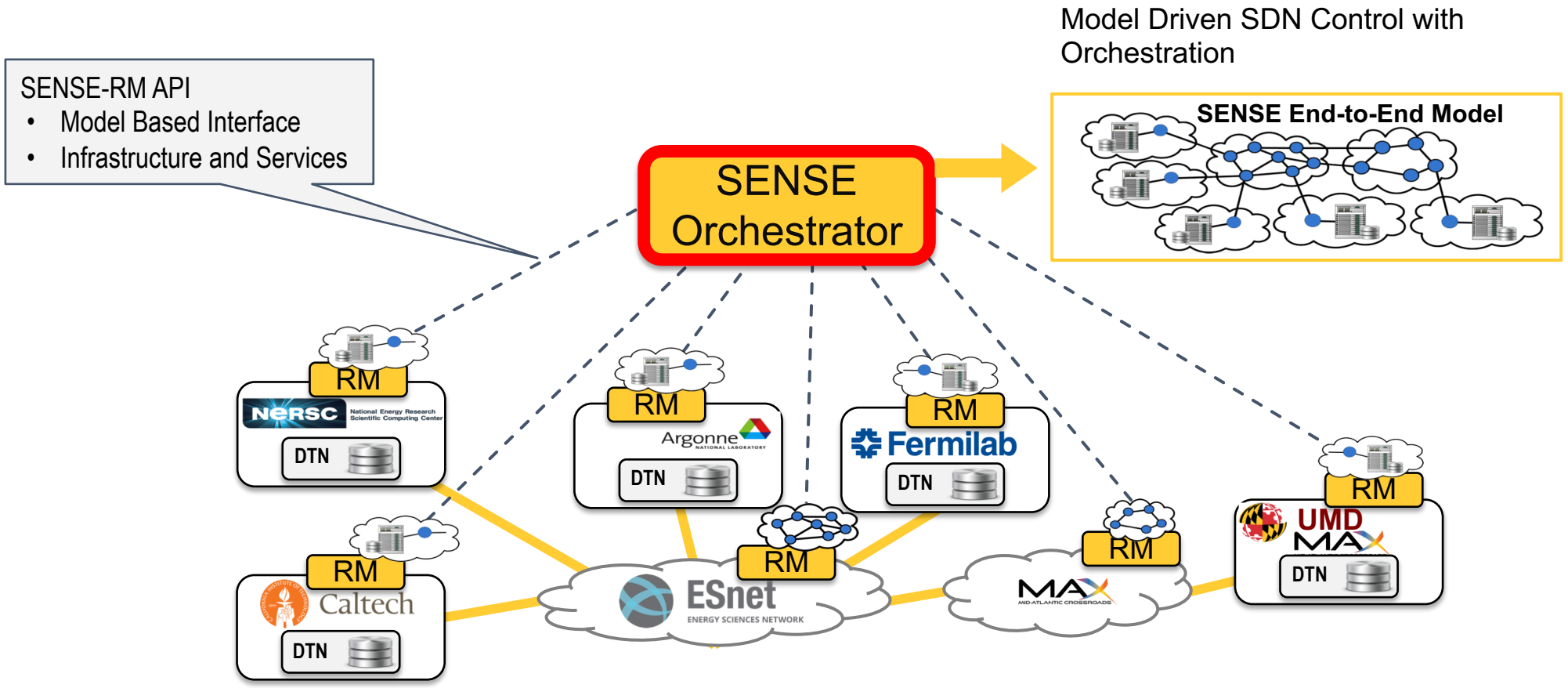
Network-RM Functions/Roles:

- Responsible for a specific set of Network Resources
- Generate real-time MRML Model
- Evaluate and respond to SENSE Orchestrator information and service requests (including negotiation)
- Provision network resources in support of SENSE services
- Provide status, monitoring, and debug functions



SENSE Service – Orchestrator

Caltech



SENSE – Filling in the gaps

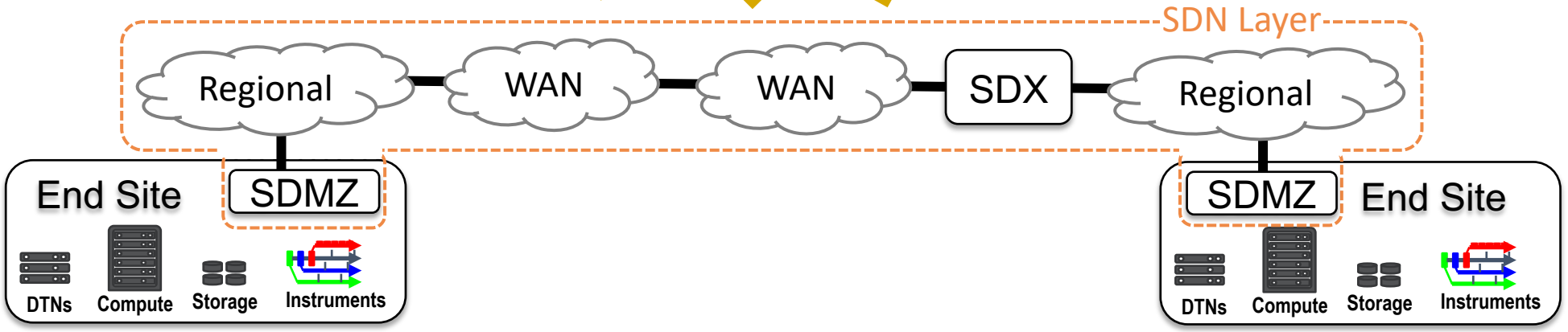
Caltech

Designed for adaption to available "SDN" systems

SENSE native "Resource Manager" available if no current automation layer is present

SENSE

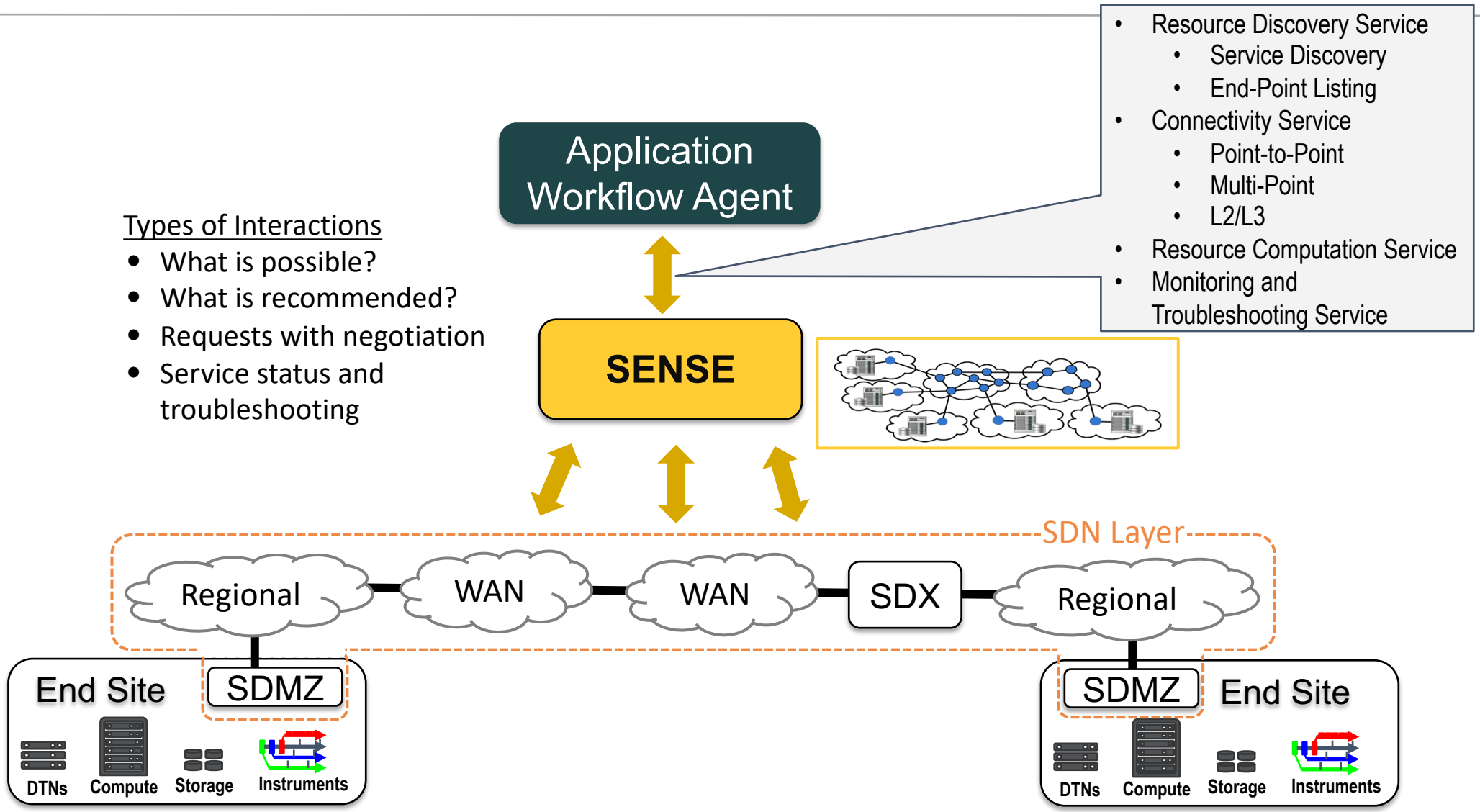
SENSE operates between the SDN Layer controlling the individual networks/end-sites



SENSE – Filling in the gaps

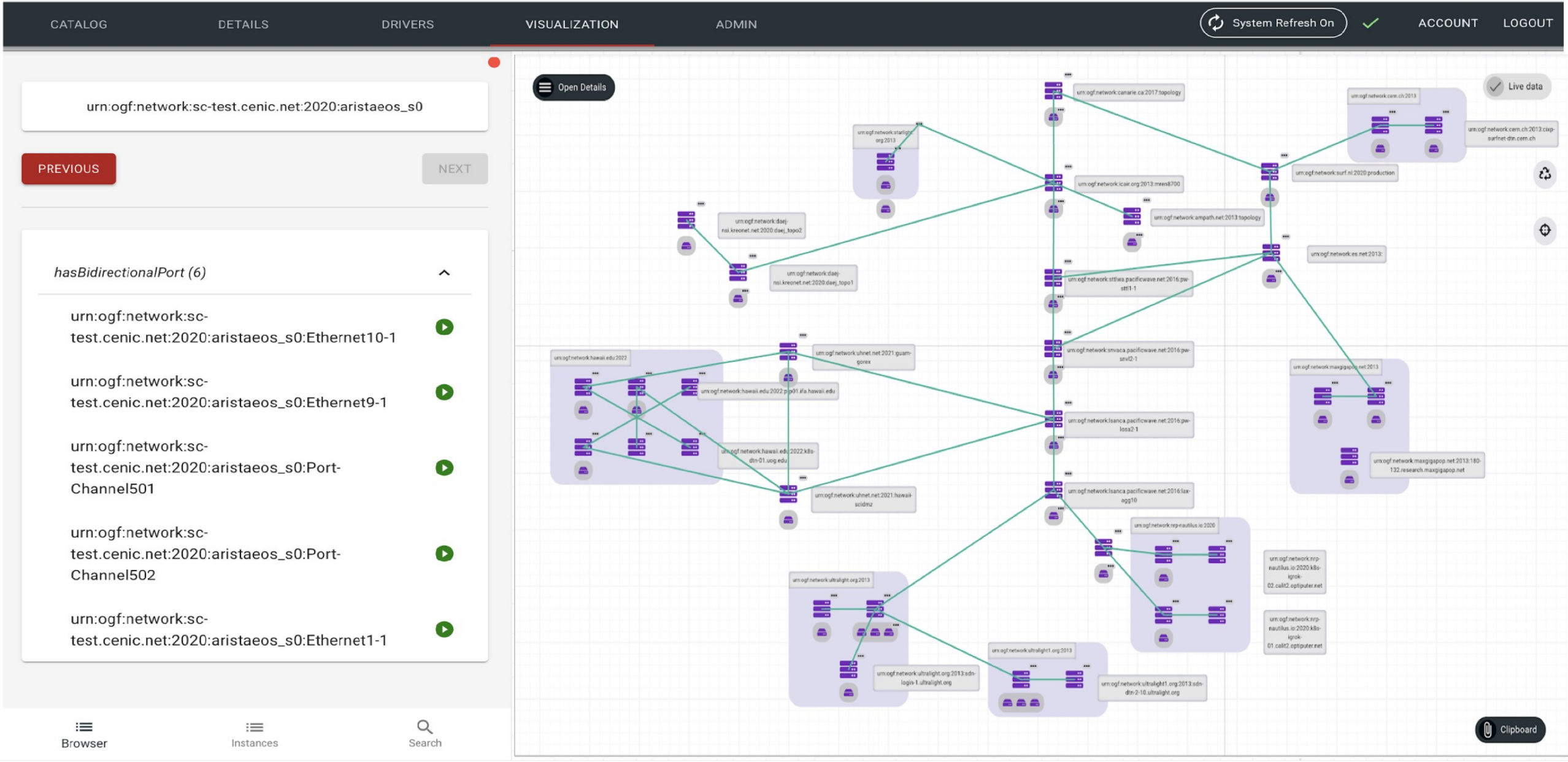
Caltech

Types of Interactions
- What is possible?
- What is recommended?
- Requests with negotiation
- Service status and troubleshooting



SENSE Model Based Resource Descriptions

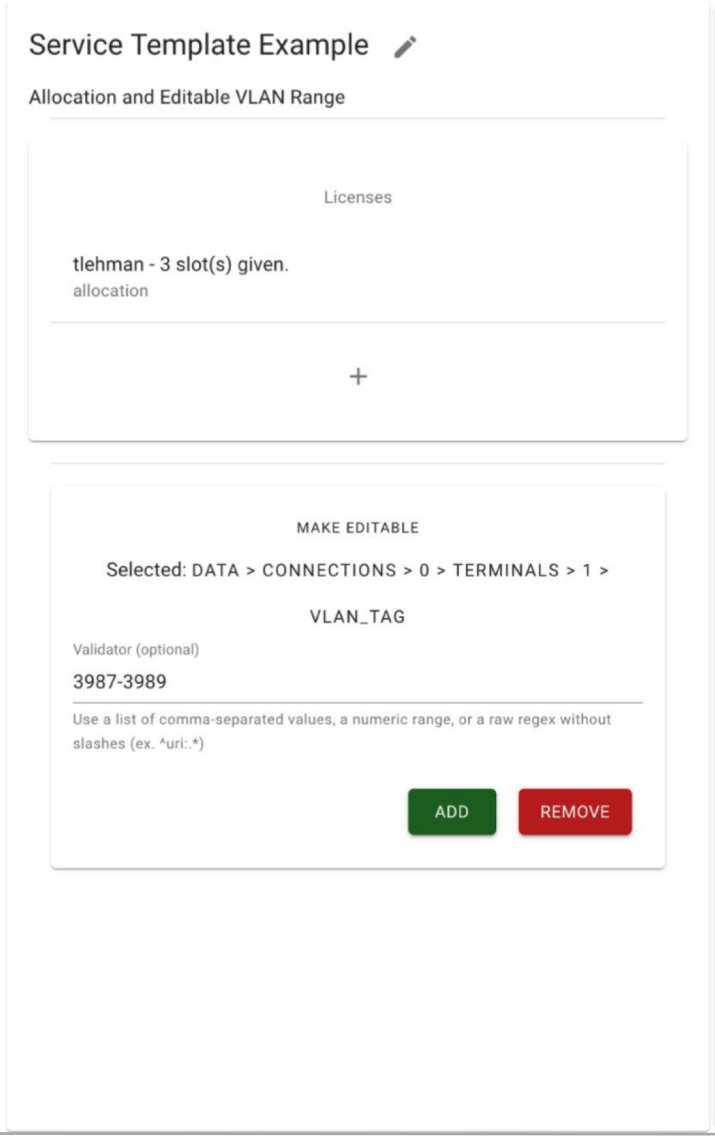
Caltech

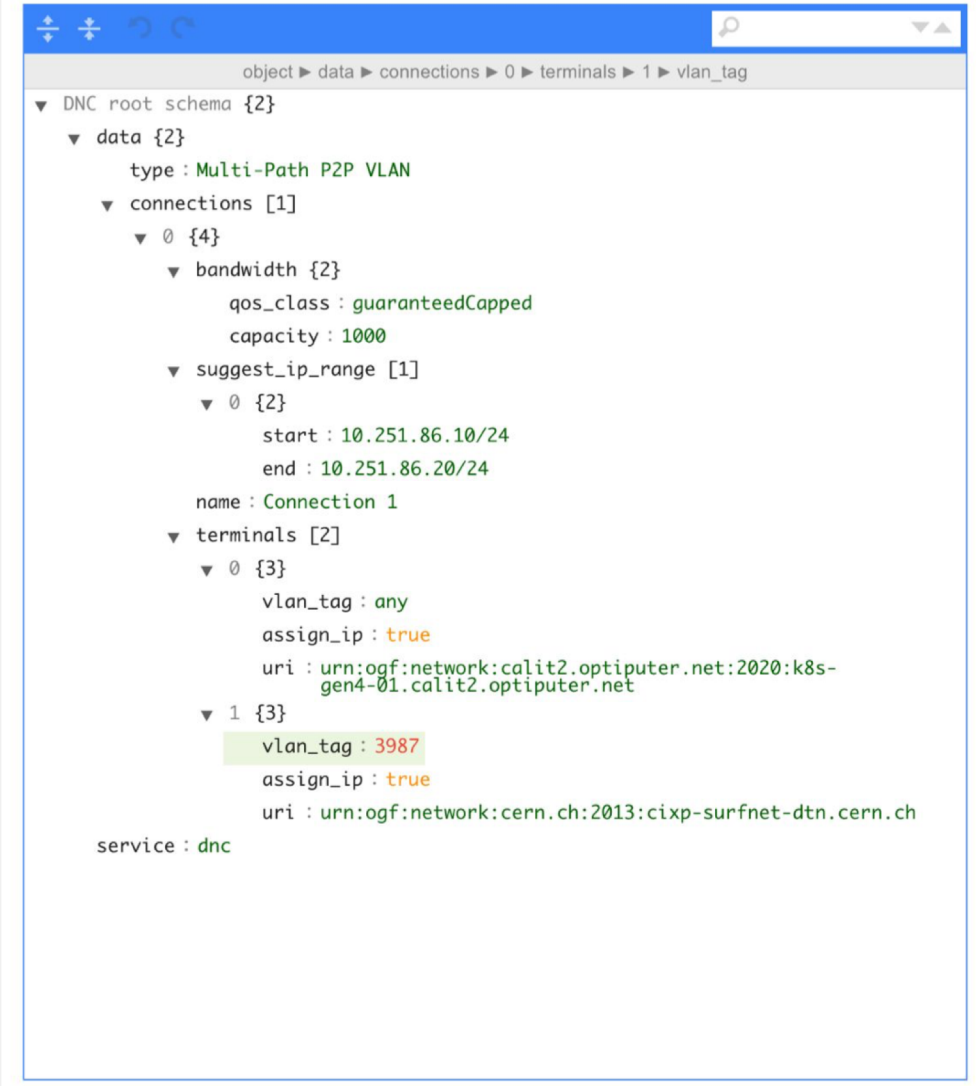


SENSE – Service Template

Caltech

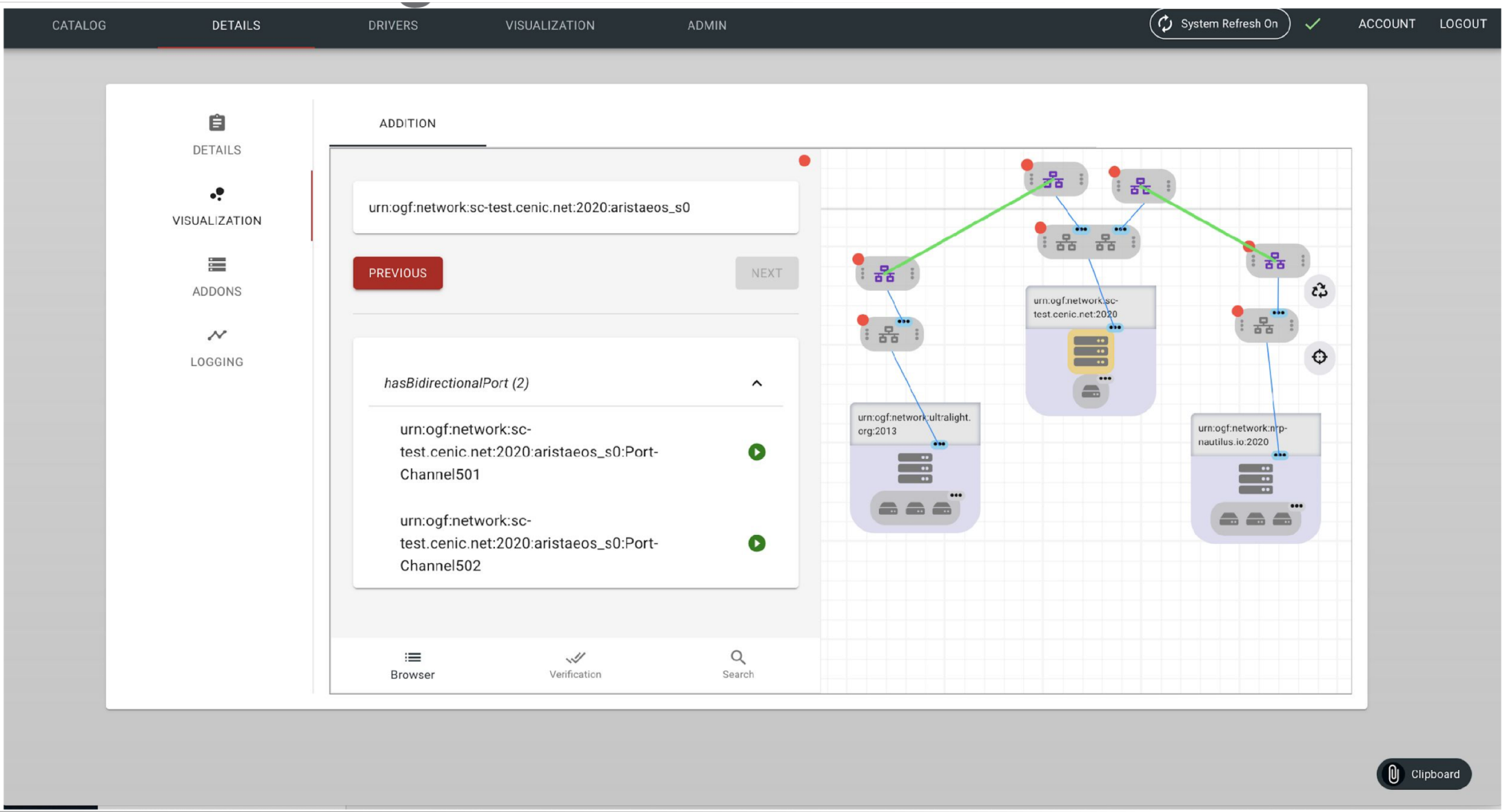
- Read only and optionally with user editable parameters
- Allows users to run with one time "ticket" or multiple time-use allocations





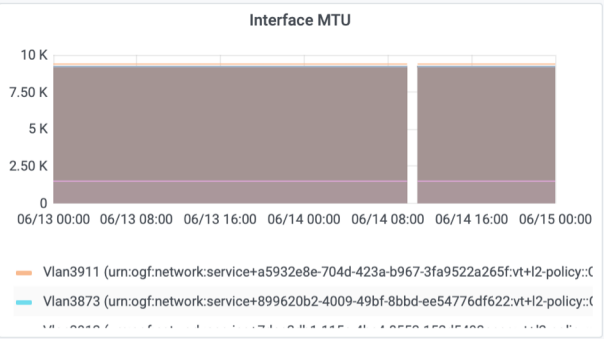
Layer 3 Flow to WAN Engineered Traffic path

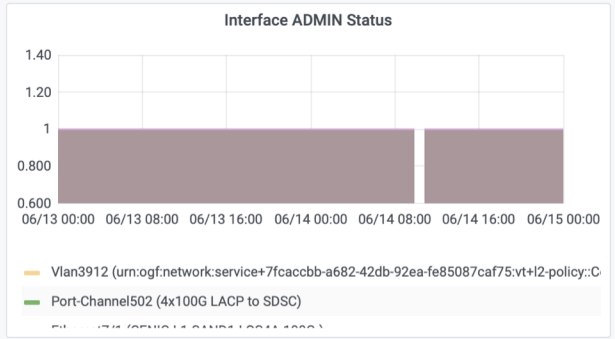
Caltech

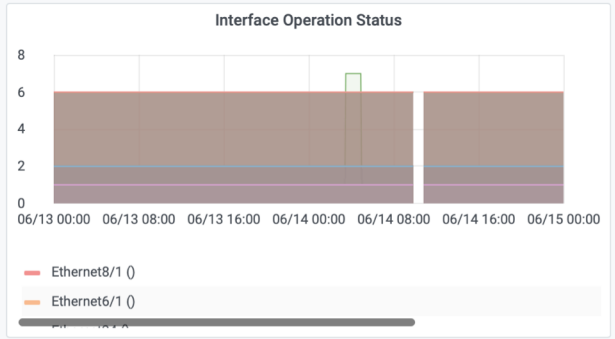


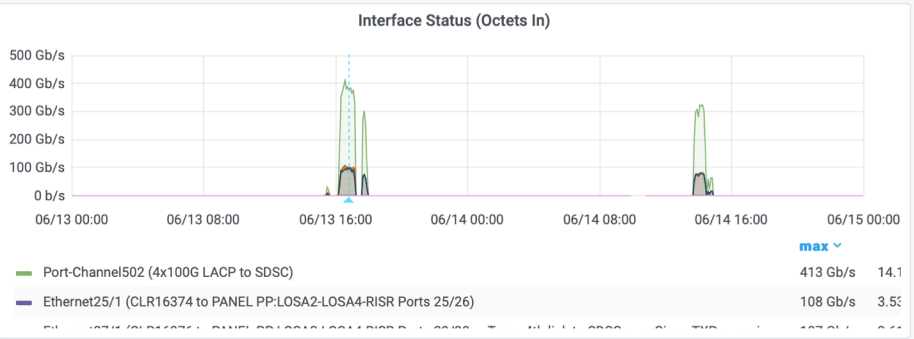
End-to-End Real Time Monitoring

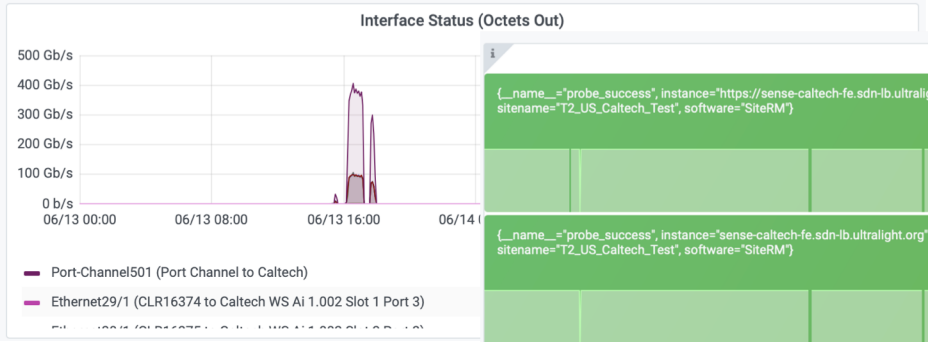
Caltech

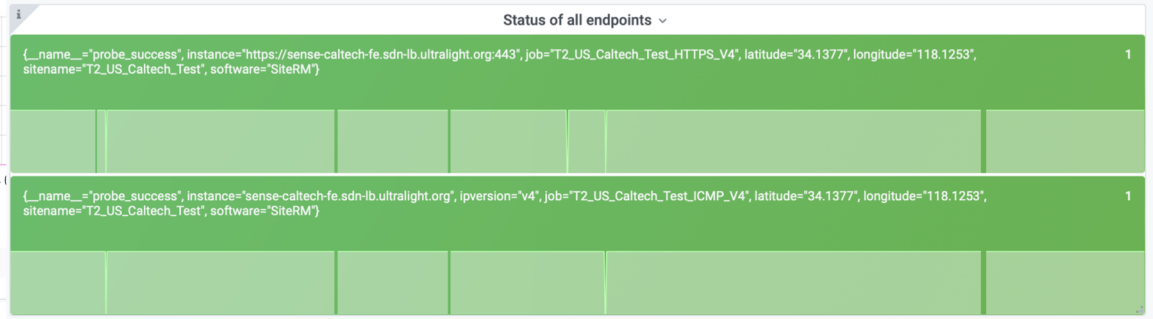


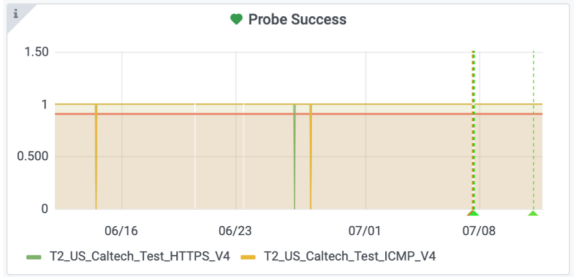


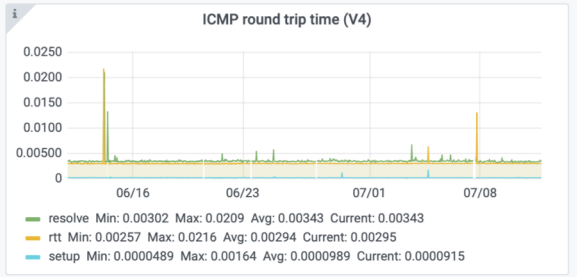


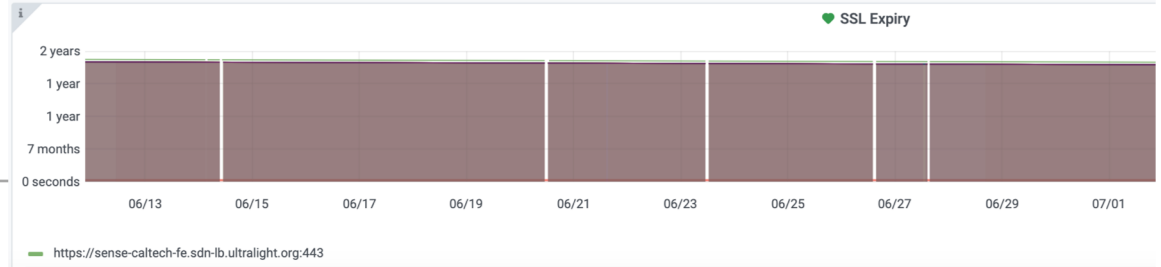












Runs on PRP Kubernetes: https://autogole-grafana.nrp-nautilus.io
Grafana, Prometheus, Blackbox Exporter, SSLExporter, Slack/Email alerting for each Site/Network RM/Orchestrator issue.

SENSE Papers and Info

Caltech

- Software-Defined Network for End-to-end Networked Science at the Exascale, Elsevier Future Generation Computer Systems, Vol 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

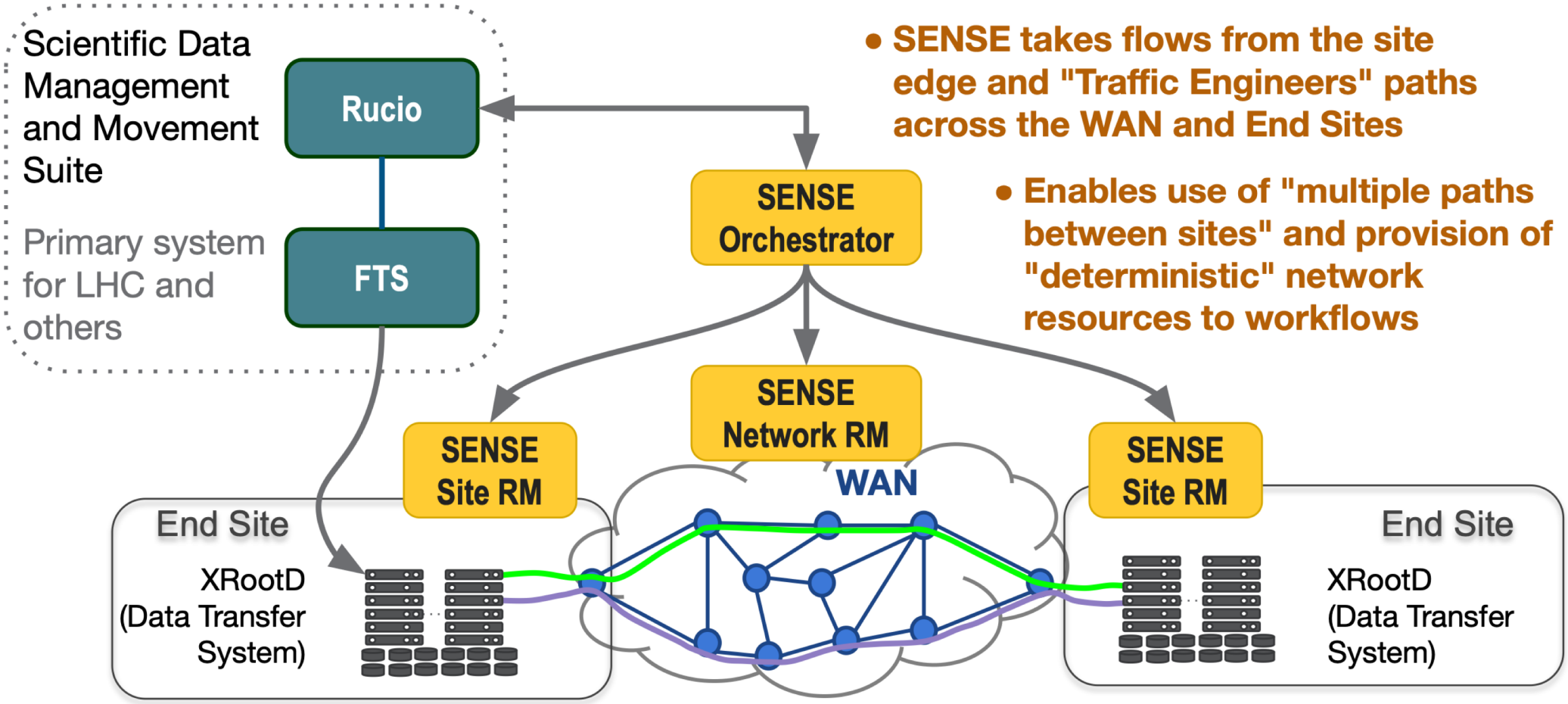
* Corresponding author.
E-mail addresses: imonga@es.net (I. Monga), chin@es.net (C. Guok), macauley@es.net (J. MacAuley), asim@lbl.gov (A. Sim), newman@hep.caltech.edu (H. Newman), jbalcas@caltech.edu (J. Balcas), demar@fnal.gov (P. DeMar), winkler@mcs.anl.gov (L. Winkler), tlehman@virnao.com (T. Lehman),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

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

Spoiler alert!

Caltech

- 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

Acknowledgments

This ongoing work is partially supported by the US National Science Foundation (NSF) Grants OAC-2030508, OAC-1841530, OAC-1836650, MPS-1148698, and PHY-1624356. In addition, the development of SENSE is supported by the US Department of Energy (DOE) Grants DE-SC0015527, DE-SC0015528, DE-SC0016585, and FP-00002494. Finally, this work would not be possible without the significant contributions of collaborators at ESnet, Fermilab, Caltech, and UCSD and many other international organizations.