SENSE
Intelligent Network Services for Science Workflows

SENSE Team

Supercomputing 2019
November 17-22, 2019

SENSE Team

ESnet ENERGY SCIENCES NETWORK
Caltech
UNIVERSITY OF MARYLAND
NERSC
MAX MID-ATLANTIC CROSSROADS
BERKELEY LAB
Argonne NATIONAL LABORATORY
Fermilab

Sponsor

U.S. DEPARTMENT OF ENERGY
Office of Science
Advanced Scientific Computing Research (ASCR)

SENSE Team

ESnet ENERGY SCIENCES NETWORK — BERKELEY LAB

- Inder Monga (PI)
- Chin Guok
- John MacAuley
- Alex Sim

Fermilab

- Phil Demar

Argonne NATIONAL LABORATORY

- Linda Winkler

NERSC

- Damian Hazen

Caltech

- Harvey Newman
- Justas Balcas
- Maria Spiropulu
- Raimondas Sirvinskas

UNIVERSITY OF MARYLAND — MAX MID-ATLANTIC CROSSROADS

- Xi Yang

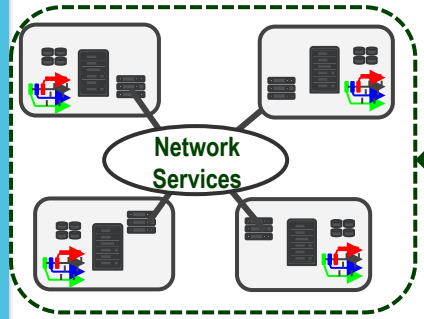
Virnao

- Tom Lehman

Vision and Objectives

- Big Science needs to coordinate (and often schedule) its utilization of distributed resources (compute, storage, instruments) in workflow specific ways.
- Distributed scientific workflows need end-to-end automation so the focus can be on science, and not infrastructure:
  - Manual provisioning and infrastructure debugging takes time
  - No service consistency across domains
  - No service visibility or automated troubleshooting across domains
  - Lack of realtime information from domains impedes development of intelligent services

Identifying the Gaps

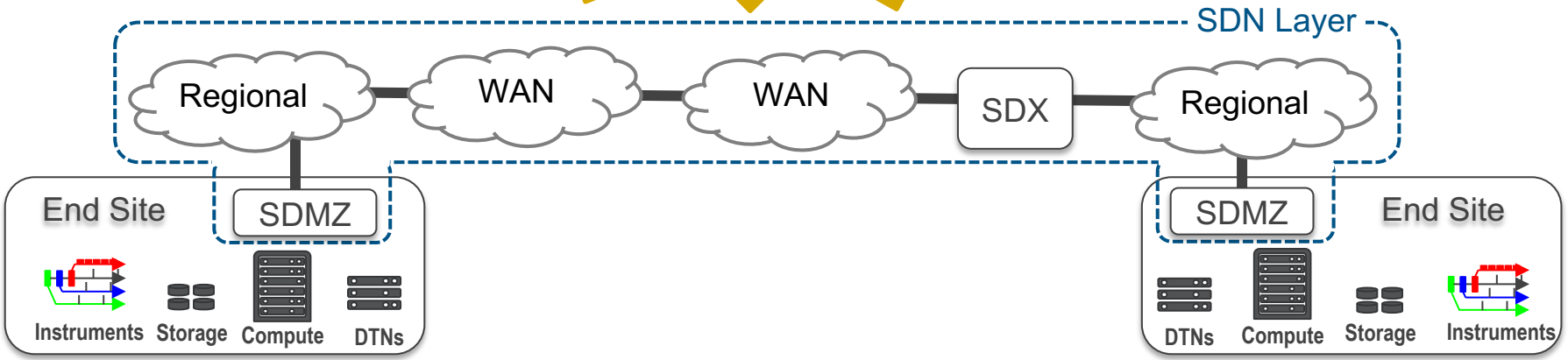


Science Workflow Specific Topology and Services

Application Workflow Agents

What goes here?

Need something to deal with Multi-Domain, Multi-Resource Distributed Infrastructure

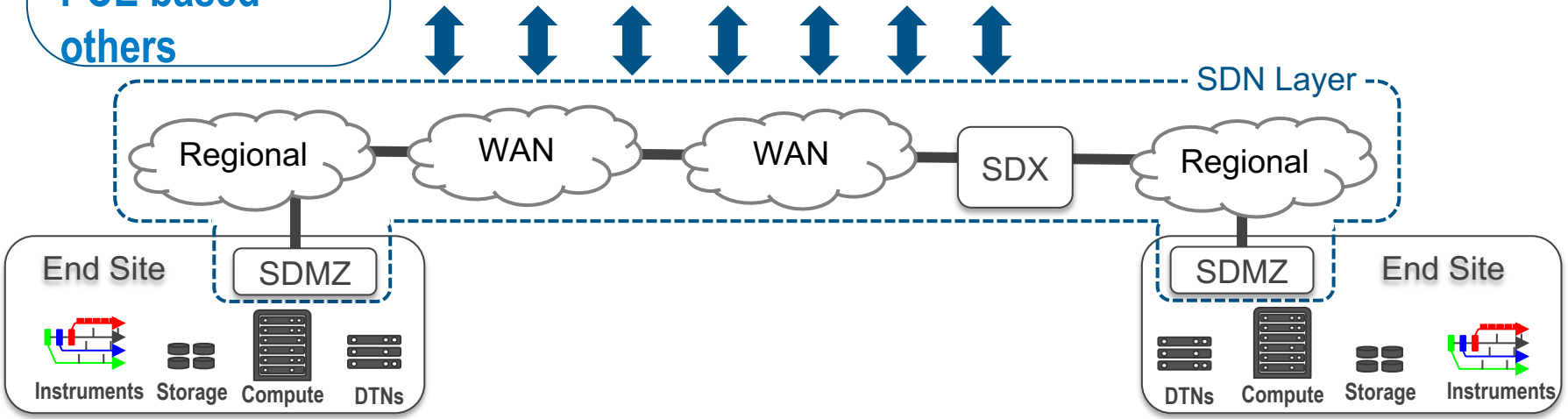


The ”SDN” Layer is Complex and Heterogenous

Control Plane
Internet2 OESS
ESnet OSCARS
OpenNSA
OpenDaylight
ONOS
PCE based
others

Data Plane/Services
Layer 1/2/3
Point-to-Point
Multi-Point
Layer 3 VPNs
QoS

Attached Distributed Resources
End Sites/SDMZs,
Clouds (Public, R&E, Edge)
HPC
Instruments
Storage/Data Lakes



SENSE - Filling in the Gaps

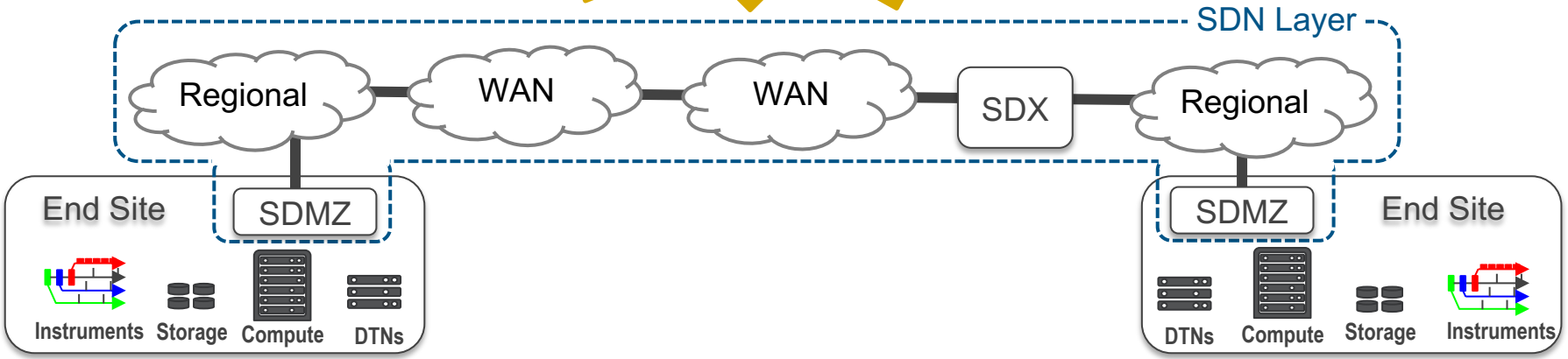
Application Workflow Agents

Designed for adaption to available “SDN” systems

SENSE

SENSE native “Resource Manager” available if no current automation layer

SENSE operates between the SDN Layer controlling the individual networks/end-sites, and science workflow agents/middleware



SENSE Solution Approach – SDN Layer Interactions

End-to-End model-based distributed resource reasoning and intelligent service orchestration

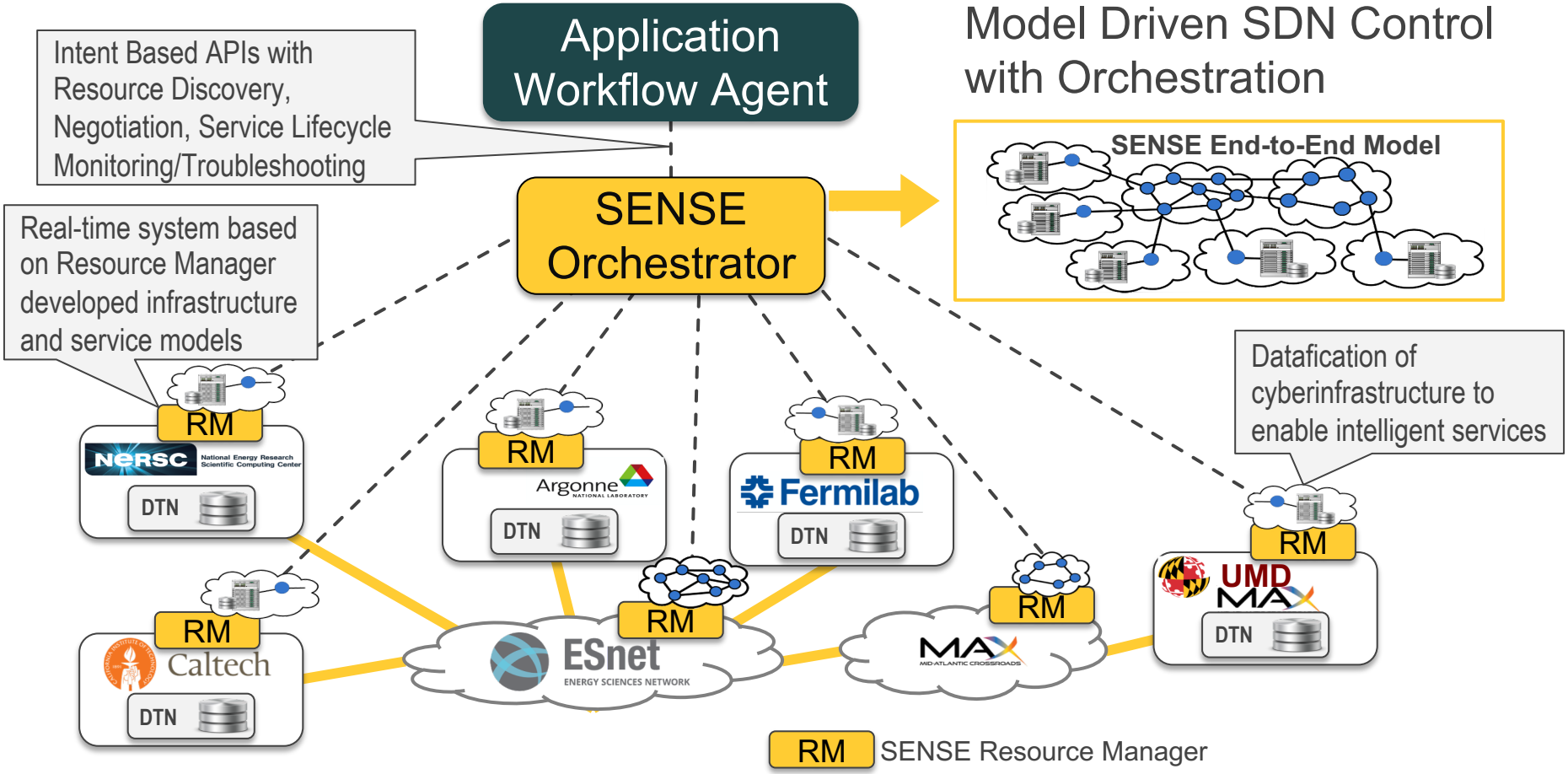
- Hierarchical service resource architecture
- Unified network and end-site resource modeling and computation
- Model based real-time control
- Application driven orchestration workflow
- End-to-end network data collection and analytics integration

SENSE Solution Approach – Application Interactions

A new paradigm for Application to Network 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.

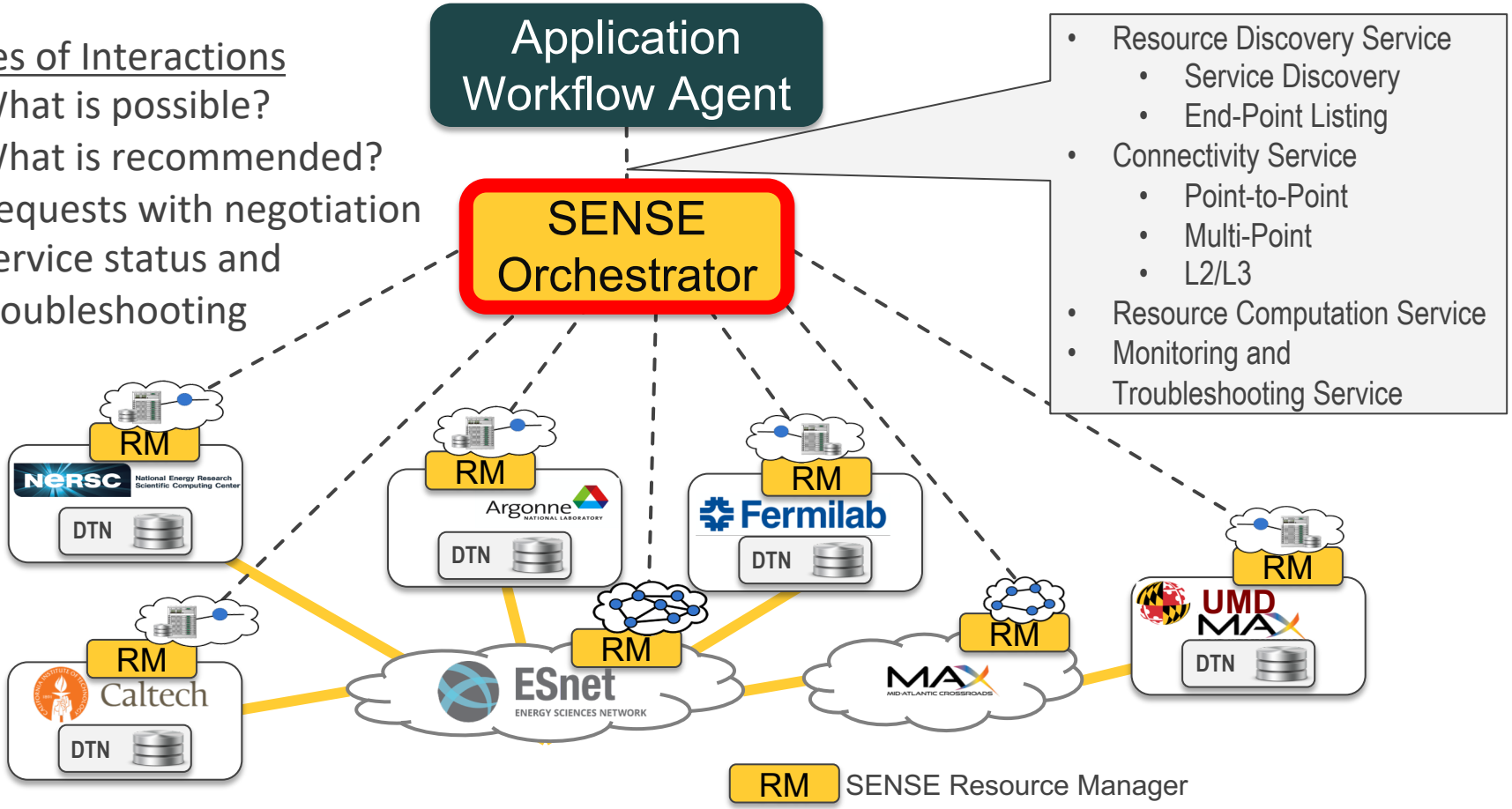
SENSE Architecture



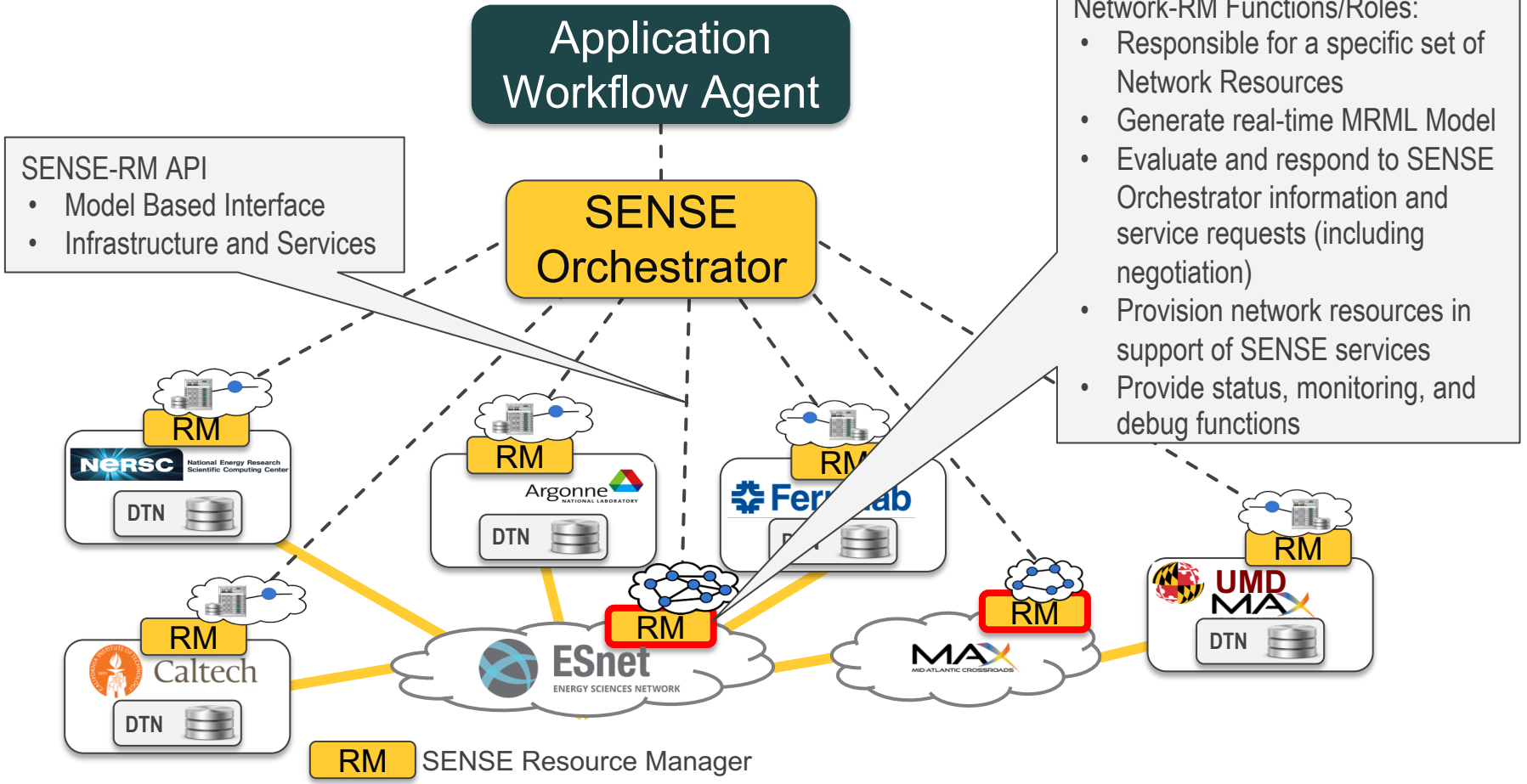
SENSE Orchestrator

Types of Interactions

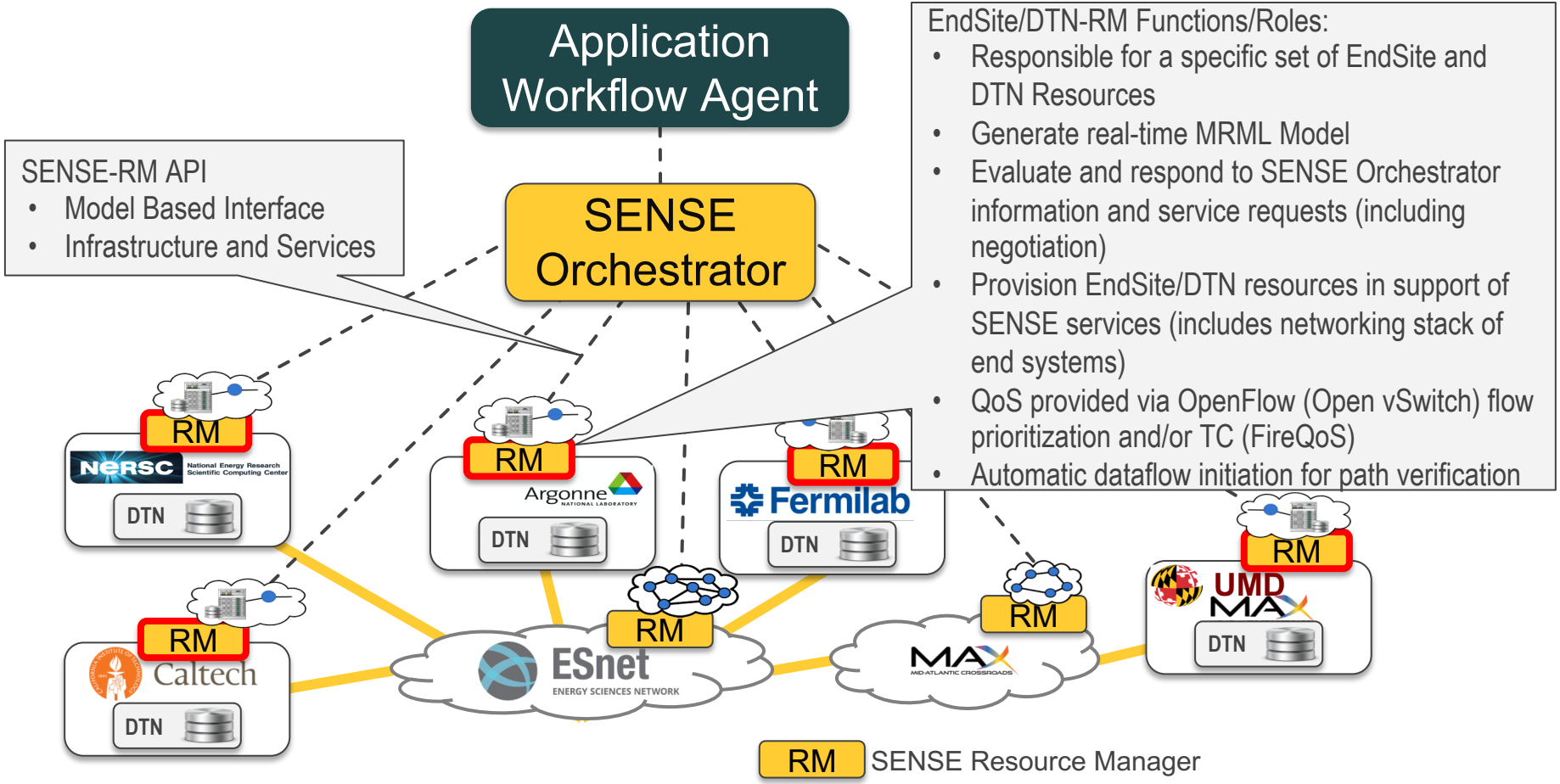
- What is possible?
- What is recommended?
- Requests with negotiation
- Service status and troubleshooting



SENSE Network Resource Manager



SENSE DTN/End-Site Resource Manager



Modeling Language and Schema

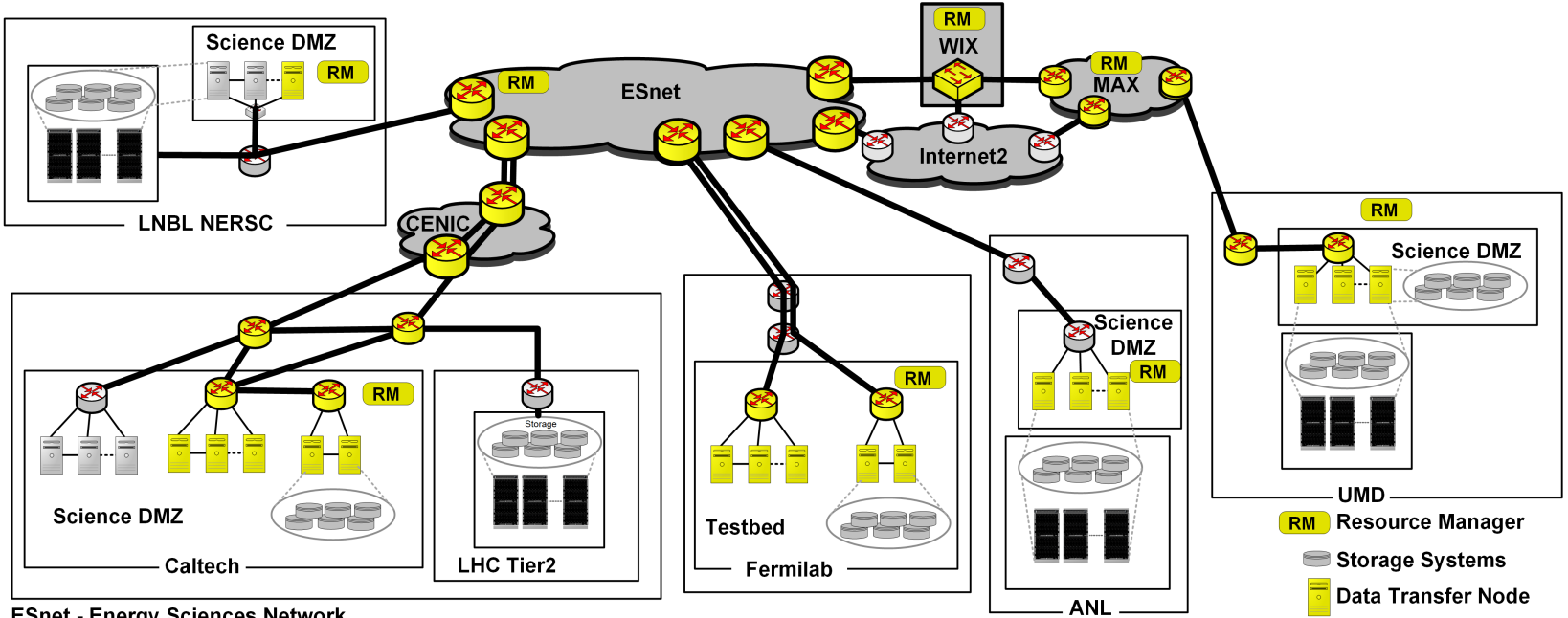
- Based on Network Markup Language (NML) standard developed by the Open Grid Forum (OGF)
  - Added extensions to allow other resource types in addition to network elements/topologies to be described and modeled

- Multi-Resource Markup Language (MRML)
  - https://github.com/esnet/nml-mrml

SENSE Data Plane Services

- **Data Plane Connectivity Services:**
  - **Point-to-Point (Layer 2)**
  - **Multi-Point (Layer 2)**
  - **Layer 3 QoS/Priority**
  - **Layer 3 VPN (provision and/or attach)**
- **Options**
  - **Layer 2 (with L3 addressing)**
  - **Layer 3 Routed Network Connections**
  - **Quality of Service (guaranteedCapped, guaranteed, bestEffort)**
  - **Negotiation**
  - **Scheduling, Batch Service Request**
  - **Strict and Loose hops, Preemption, Lifecycle monitoring and debug**

SENSE Testbed



ESnet - Energy Sciences Network
LBNL NERSC - Lawrence Berkeley National Laboratory National Energy Research Computing Center
Fermilab - Fermi National Accelerator Laboratory
ANL - Argonne National Laboratory
Caltech - California Institute of Technology
Internet2 - University Consortium Wide Area Network Internet2
CENIC - California Research and Education Network
UMD - University of Maryland
MAX - Mid-Atlantic Crossroads

Color indicate device is under SENSE control
Color indicate device is not under SENSE control

SENSE SC19 L3VPN Demonstration

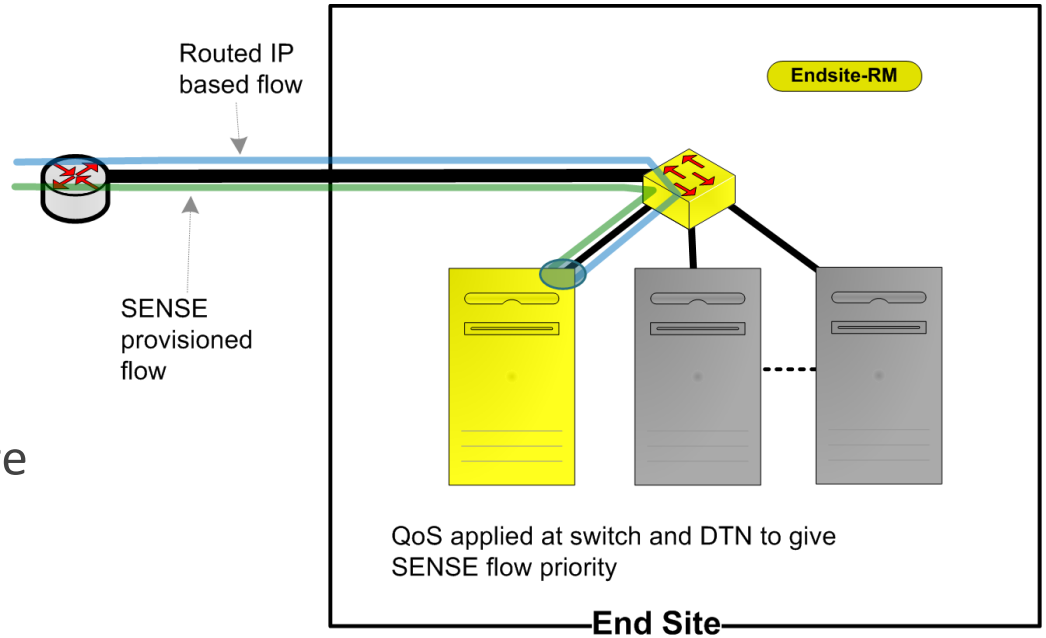
- SENSE Layer 3 service which provides the mechanisms for directing desired traffic onto specific Layer 3 Virtual Private Networks for policy and/or quality of service reasons.
  - L3-VPN-Provision - New L3-VPN provision
  - L3-VPN-Attach - Attachment to an existing L3-VPN. May include the automated establishment of a BGP peering, or may be a simpler attachment based on a default route and/or static routes being added to the attaching system.
  - Allows workflow middleware to redirect traffic at granularities ranging from a single flow, specific end-system, or an entire end-site onto the desired SENSE provisioned service.
  - Dynamic switch between L2 pt-to-pt and L3 VPN (Auto re-ip addressing)

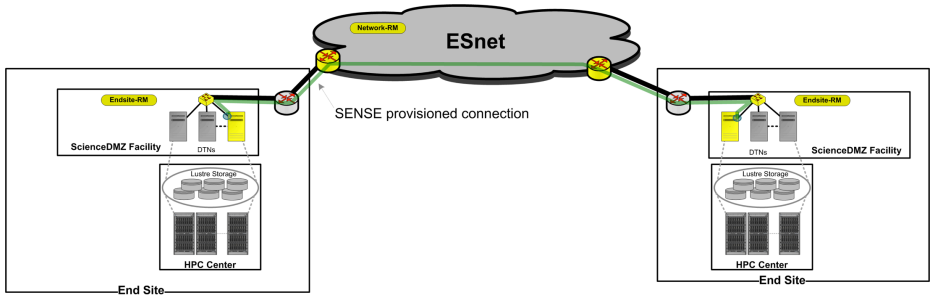
Use Cases

- **Data Transfer Node Priority Flow (SENSE Enabled DTNs)**
  - Deterministic end-to-end data transfers
- **DOE Superfacility**
- **Exascale for Free Electron Lasers (ExaFEL)**
  - Streaming the data from the LCLS (SLAC) online cache (NVRAM) to the NERSC data transfer nodes
- **Large Hadron Collider/ Compact Muon Solenoid (LHC/CMS) File Transfer Service (FTS)**
  - Use of SENSE paths based on transfer queues and sizes
- **Big Data Express**
  - Intelligent selection of WAN paths based on user requirements

SENSE Enabled DTNs

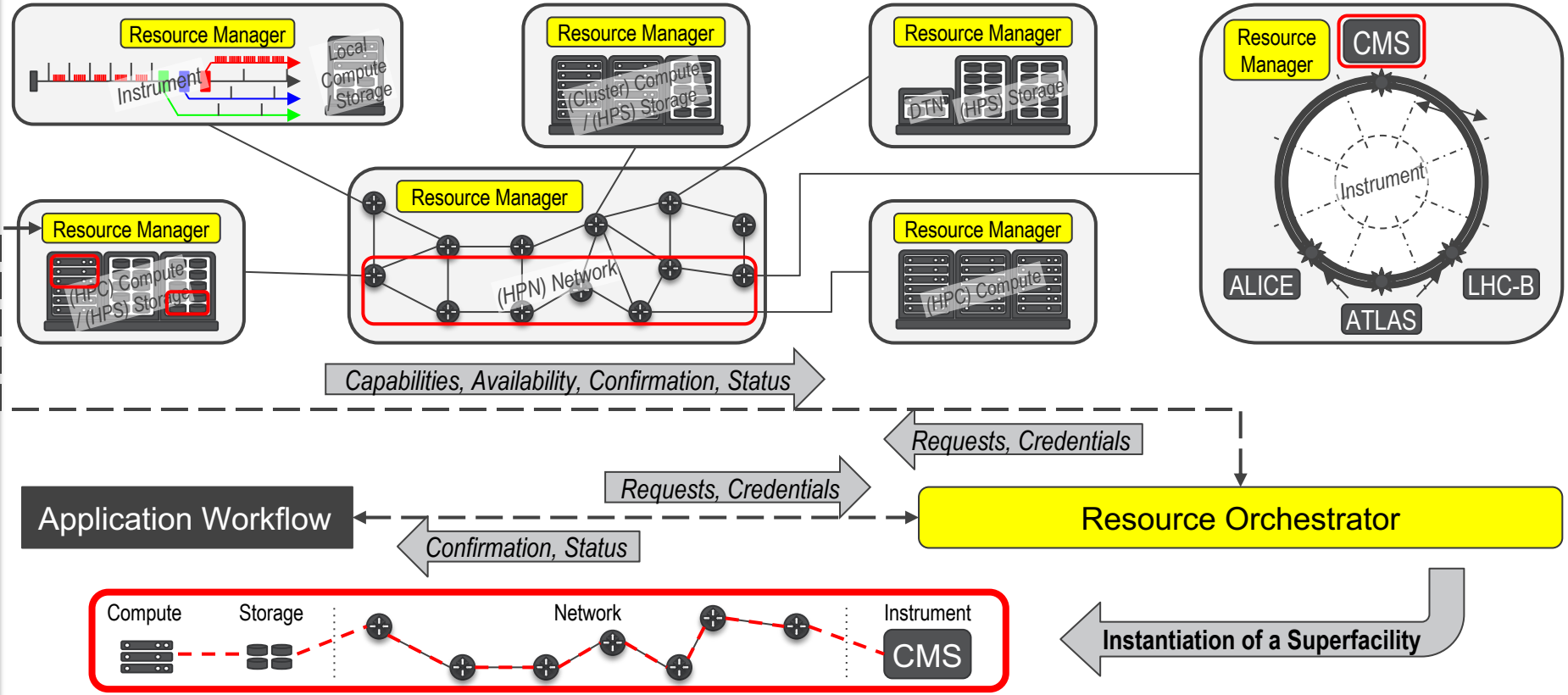
- SENSE DTNs can be deployed next to production DTNs
- No impact to standard DTN operations
- Just adds a “priority flow” feature



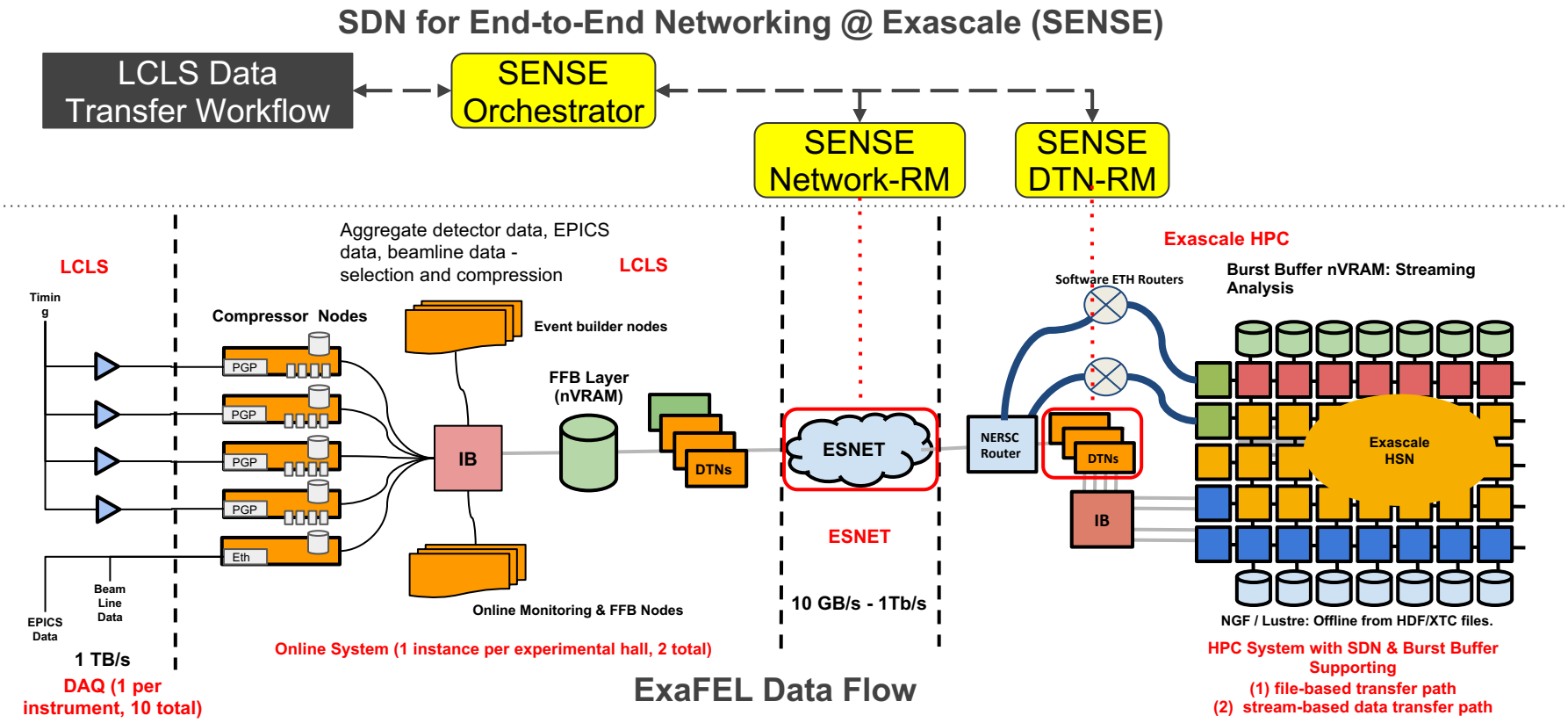


- Scheduled and guaranteed resources, network and end system
- Can be included as part of application workflow planning

Superfacility Automation

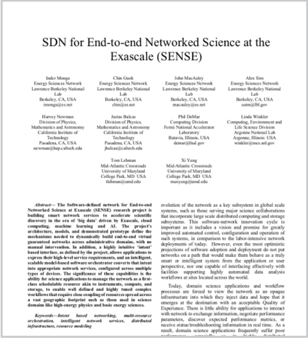


ExaFEL Use Case - Superfacility Automation Prototyping



2018 INDIS Workshop at SC18 - SENSE Paper, Presentation, and Demos

- 2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS)
  - SDN for End-to-end Networked Science at the Exascale (SENSE) Paper
  - http://sense.es.net/publications



SDN for End-to-end Networked Science at the Exascale (SENSE)

Inder Monga
Energy Sciences Network
Lawrence Berkeley National Lab
Berkeley, CA, USA
imonga@es.net

Chin Guok
Energy Sciences Network
Lawrence Berkeley National Lab
Berkeley, CA, USA
chin@es.net

John MacAuley
Energy Sciences Network
Lawrence Berkeley National Lab
Berkeley, CA, USA
macauley@es.net

Alex Sim
Energy Sciences Network
Lawrence Berkeley National Lab
Berkeley, CA, USA
asim@lbl.gov

Harvey Newman
Division of Physics, Mathematics and Astronomy
California Institute of Technology
Pasadena, CA, USA
newman@hep.caltech.edu

Justas Balcas
Division of Physics, Mathematics and Astronomy
California Institute of Technology
Pasadena, CA, USA
jbalcas@caltech.edu

Phil DeMar
Computing Division
Fermi National Accelerator Laboratory
Batavia, Illinois, USA
demar@fnal.gov

Linda Winkler
Computing, Environment and Life Science Division
Argonne National Lab
Argonne, Illinois USA
winkler@mcs.anl.gov

Tom Lehman
Mid-Atlantic Crossroads
University of Maryland
College Park, MD USA
tlehman@umd.edu

Xi Yang
Mid-Atlantic Crossroads
University of Maryland
College Park, MD USA
maxyang@umd.edu

Abstract— **The Software-defined network for End-to-end Networked Science at Exascale (SENSE) research project is building smart network services to accelerate scientific discovery in the era of 'big data' driven by Exascale, cloud computing, machine learning and AI. The project's architecture, models, and demonstrated prototype define the mechanisms needed to dynamically build end-to-end virtual guaranteed networks across administrative domains, with no manual intervention. In addition, a highly intuitive 'intent' based interface, as defined by the project, allows applications to express their high-level service requirements, and an intelligent, scalable model-based software orchestrator converts that intent into appropriate network services, configured across multiple types of devices. The significance of these capabilities is the ability for science applications to manage the network as a first-class schedulable resource akin to instruments, compute, and storage, to enable well defined and highly tuned complex workflows that require close coupling of resources spread across a vast geographic footprint such as those used in science domains like high-energy physics and basic energy sciences.**

Keywords—Intent based networking, multi-resource orchestration, intelligent network services, distributed infrastructure, resource modeling

evolution of the network as a key subsystem in global scale systems, such as those serving 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 greatly improved automated control, configuration and operation of such systems, in comparison 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 located across 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 applications to interact with network to exchange information, negotiate performance parameters, discover expected performance metrics, or receive status/troubleshooting information in real time. As a result, domain science applications frequently suffer poor

- Additional SENSE Information, Demonstration movies available here:
  - http://sense.es.net

SENSE SC19 Presentation and Demonstration

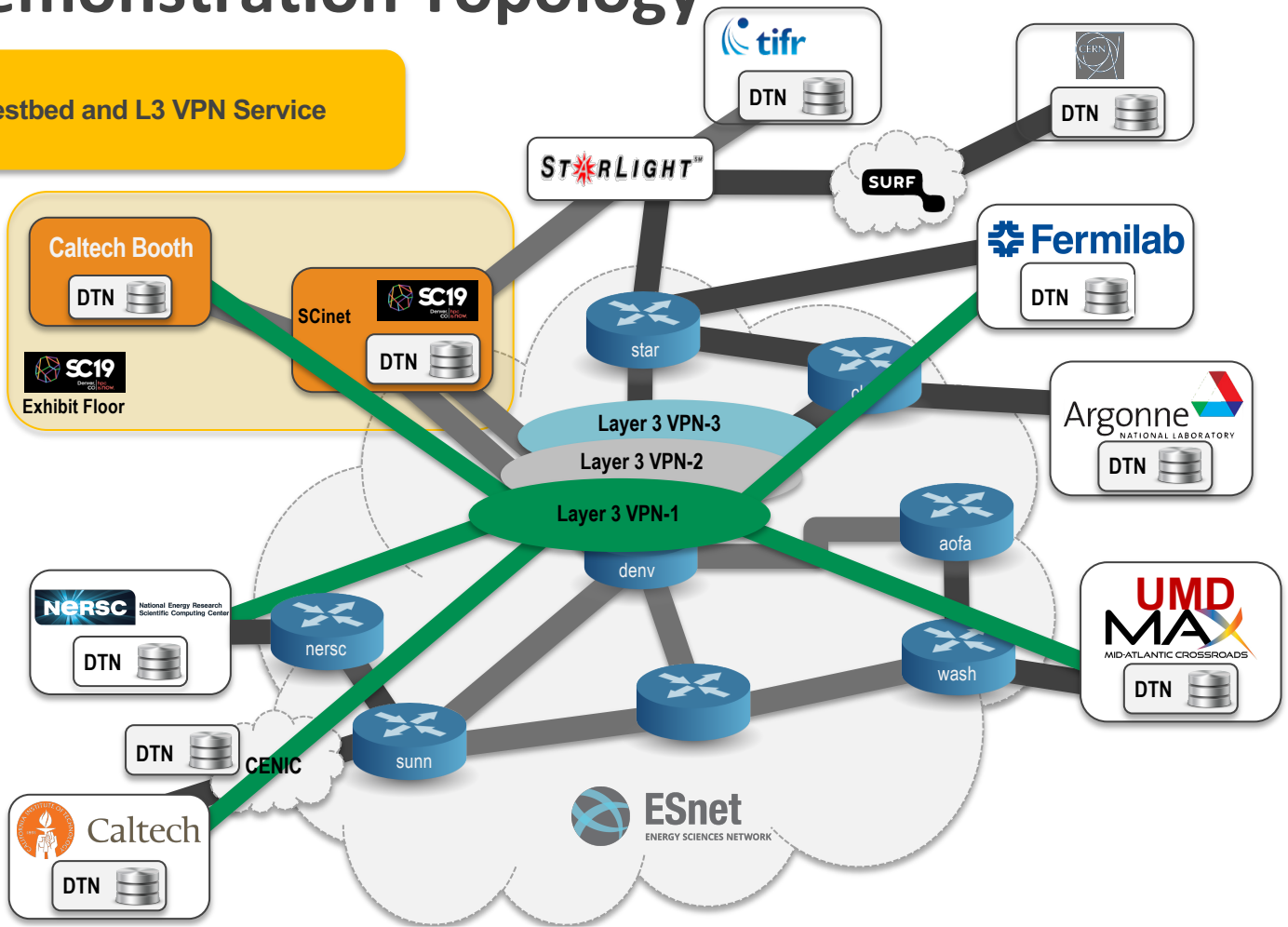
- Schedule
  - Location: Booth 543 (California Institute of Technology /CACR)
  - Schedule:
    - Wednesday
    - November 20, 2019
    - 11am and 4pm
- SENSE: Intelligent Network Services for Science Workflows
  - Part of the SC19 Network Research Exhibition Demonstration
  - https://sc19.supercomputing.org/scinet/network-research-exhibition/ → SC19-NRE-013

SENSE SC19 Demonstration Topology

SENSE Testbed and L3 VPN Service

SENSE enabled resources at DOE Laboratories, Universities, Research Facilities, and SC19

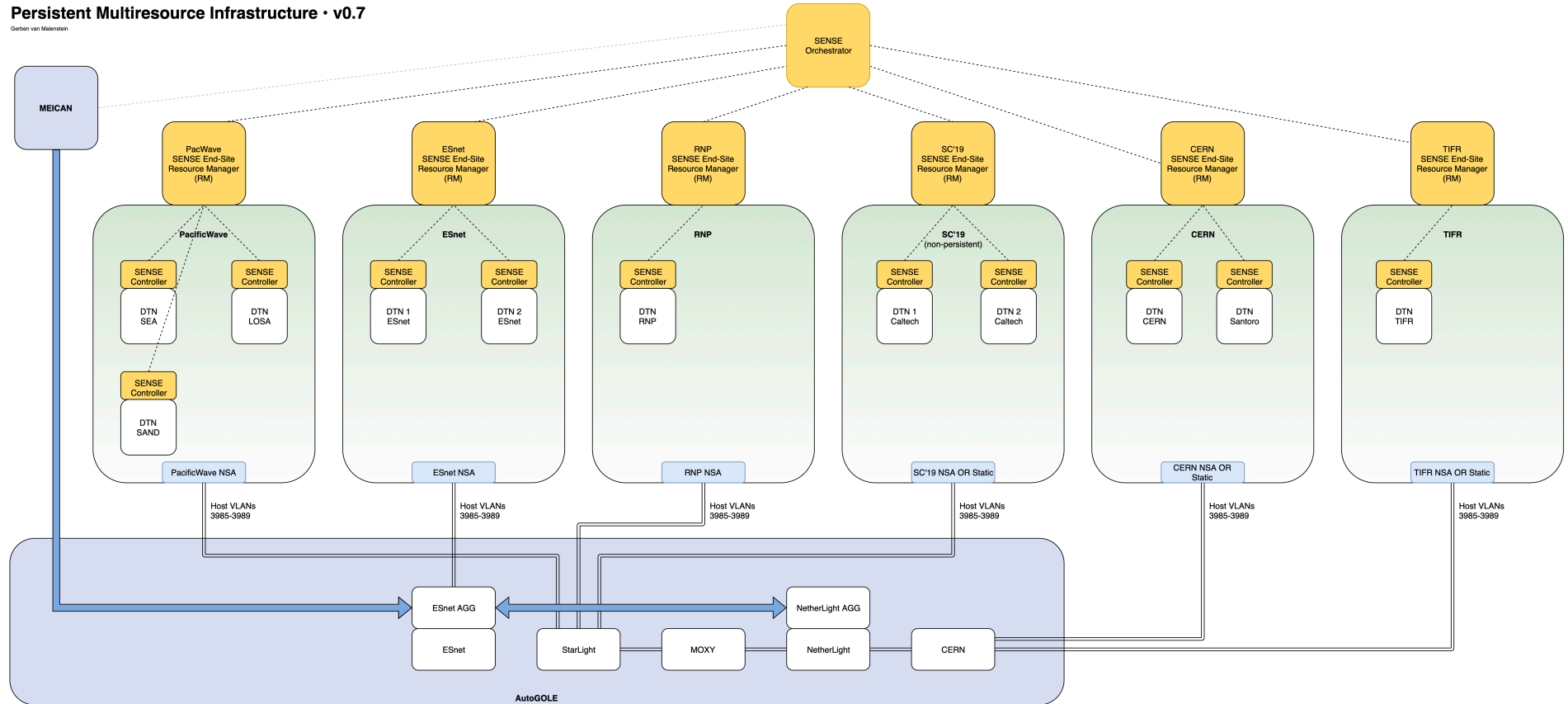
Dynamic attachment of End Site resources to L3VPNs advertised by ESnet



Questions...

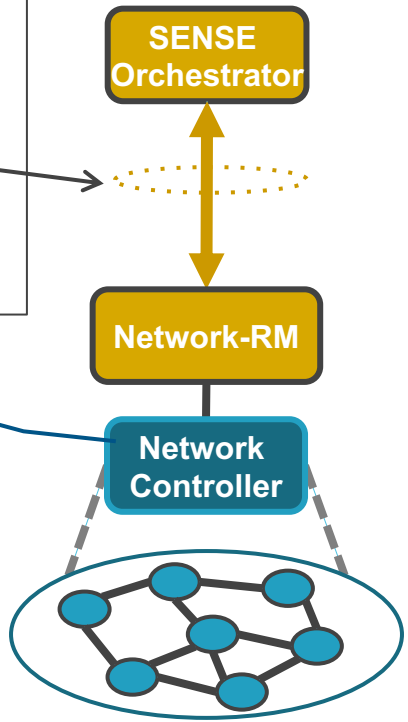
Extra Slides

SC19-NRE-020 Demonstration – Multi-Resource Orchestration via AutoGole and SENSE



SENSE Network Resource Manager (RM)

- SENSE-RM API
- Model Based Interface
- Infrastructure and Services



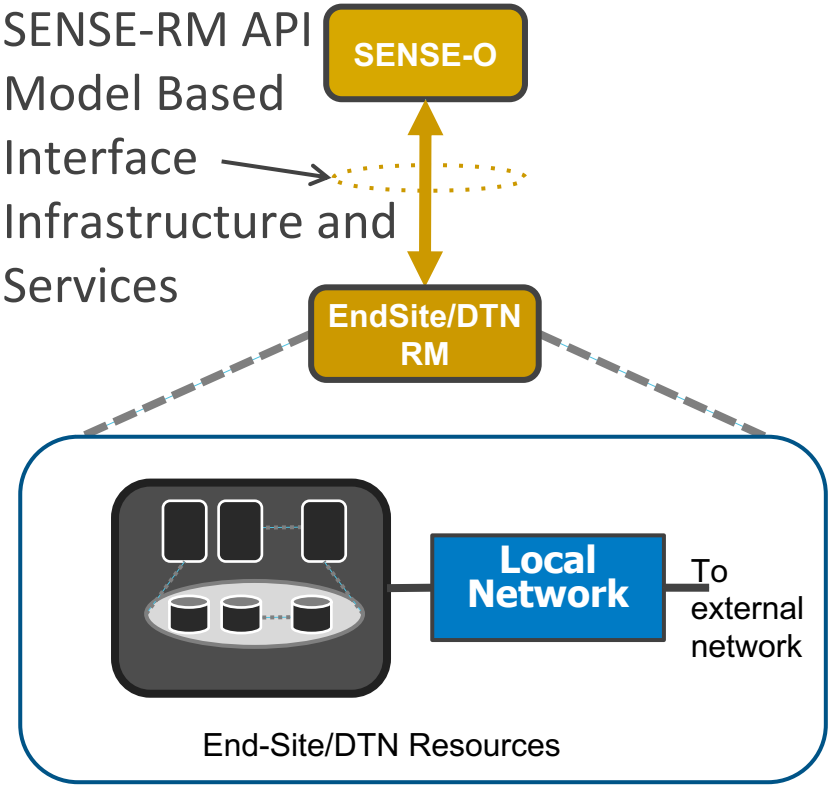
OSCARS/NSI
SDN Controller
Others

Network-RM Functions/Roles:

- **Responsible for a specific set of Network Resources**
- **Generate realtime 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 DTN/End Site Resource Manager

- SENSE-RM API
- Model Based Interface
- Infrastructure and Services



EndSite/DTN-RM Functions/Roles:

- **Responsible for a specific set of EndSite and DTN Resources**
- **Generate realtime MRML Model**
- **Evaluate and respond to SENSE Orchestrator information and service requests (including negotiation)**
- **Provision EndSite/DTN resources in support of SENSE services (includes networking stack of end systems)**
- **QoS provided via OpenFlow (Open vSwitch) flow prioritization and/or TC (FireQoS)**
- **Automatic dataflow initiation for path verification**

Application Workflow Agent Services Examples

- **Time-Block-Maximum Bandwidth (TBMB):** Application asks for a specific time block and would like to know (or provision) the maximum bandwidth available for a specific time period.
- **Bandwidth-Sliding-Window (BSW):** Application asks for a specific bandwidth and duration and provides an acceptable time window. For example, a request may be for 40 Gbps for a 10-hour time window, sometime in the next 3 days.
- **Time-Bandwidth-Product (TBP):** Application asks for “8 hours of transfer at 10Gbps” representing a TBP of 36 TBytes. The user also specifies an acceptable time window, and other options such as “prefer the highest bandwidth rate available”, or the lowest.

Application Workflow Agent Services Interactions

- **Immediate Provision:** If SENSE finds a resource path which satisfies the application request, provisioning starts immediately (after routine confirmations from both sides).
- **What is Possible?:** In this mode, SENSE simply conducts a “Resource Computation” and provides the results back to the requestor. No provisioning action is taken without further explicit requests from the user.
- **Negotiation:** One or more rounds of Resource Computation requests with subsequent provisioning request by the application user if desired.

SENSE Control Feedback Loop

Model Based Control and Orchestration

SENSE Orchestrator

Allow the machines to automate, iterate, react, and adjust to find solutions and not bring the humans in until absolutely necessary