MAXedge Services

Xi Yang, Interim Director of Research and Development

Kevin Hildebrand, HPC Architect

Alberto Jimenez, Software Engineer

Deepthought2

- Built in 2014
- 488 nodes, 9840 cores, E5-2680v2 (2.8GHz)
- 128GB of RAM each (6.4GB/core)
- 40 GPU nodes (2 Nvidia K40 GPUs each)
- 5 Large memory nodes (1TB RAM each)
- FDR (56Gbps) Infiniband
- 1PB Lustre

MARCC/Bluecrab

- Built in 2014, some more recent additions
- UMD owns 15% share
- 846 nodes, 21792 cores (most Haswell 2.5GHz), additional Broadwell, Ivy Bridge and Skylake (2.5GHz-3GHz)
- 128 GB of RAM each
- 50 GPU nodes with 2 Nvidia K80s each
- Handful of GPU nodes with Nvidia P100s
- 50 Large memory nodes (1TB RAM each)
- FDR (56 Gbps) Infiniband
- 2PB Lustre, 20PB ZFS/NFS

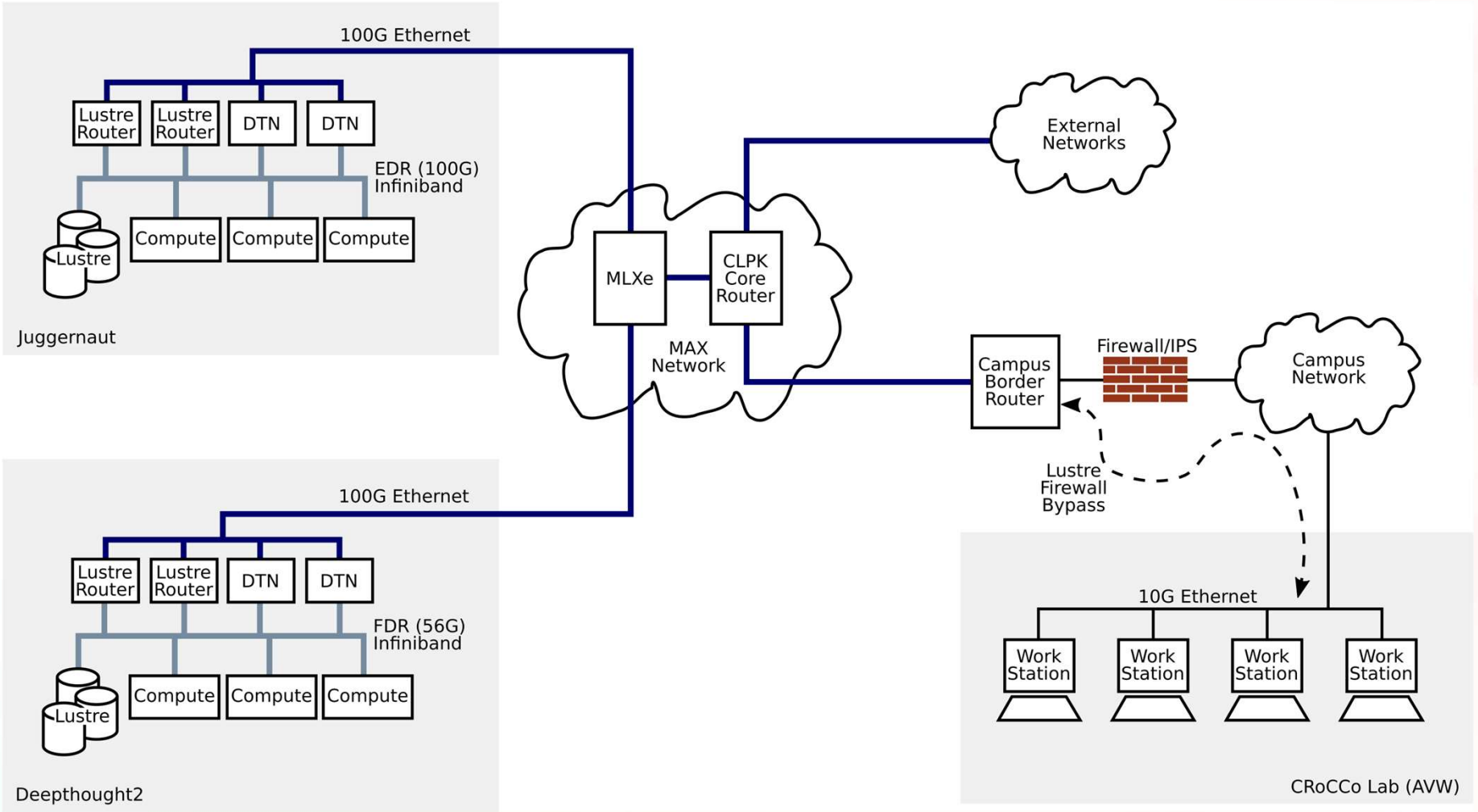
Juggernaut

- Built in November, 2018
- 15 nodes, 600 cores (Skylake 2.4GHz)
- 384GB of RAM each (9.6 GB/core)
- 1 GPU node (with 2 Nvidia P100s)
- EDR (100Gbps) Infiniband
- 1.5 PB Lustre

Deepthought3

- Currently in the planning stages
- Somewhere in the 2 Petaflop range
- Will have a small number (<10) of large memory nodes (1.5TB RAM)
- Will have around 40 GPUs, probably Nvidia V100
- Lustre storage (~2PB)
- Additional archival storage
- HDR/EDR (200Gbps/100Gbps) Infiniband

UMD/MAX HPC Networking



Outline

- What is MAXedge?
- How do we use MAXedge?
- What services does MAXedge offer?
- How does MAXedge work? – A deep dive

What is MAXedge?

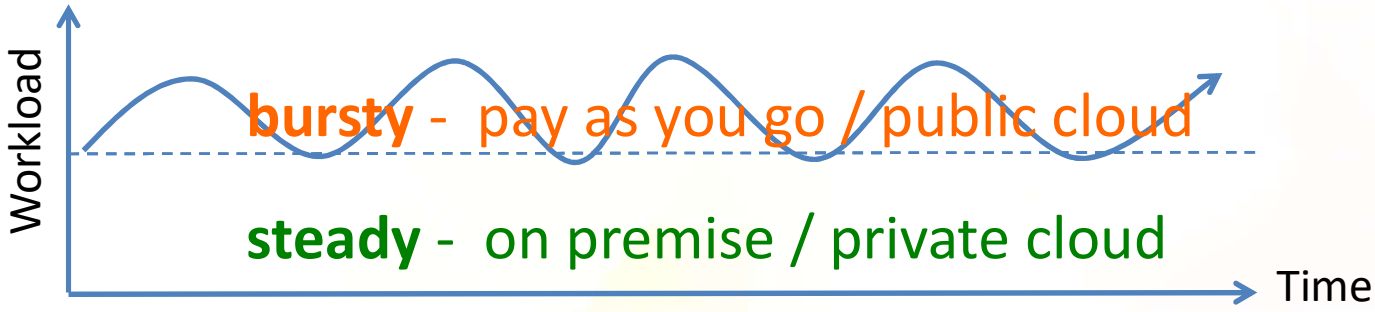
MAXedge is an intelligent, advanced and innovative edge computing facility operated by MAX.

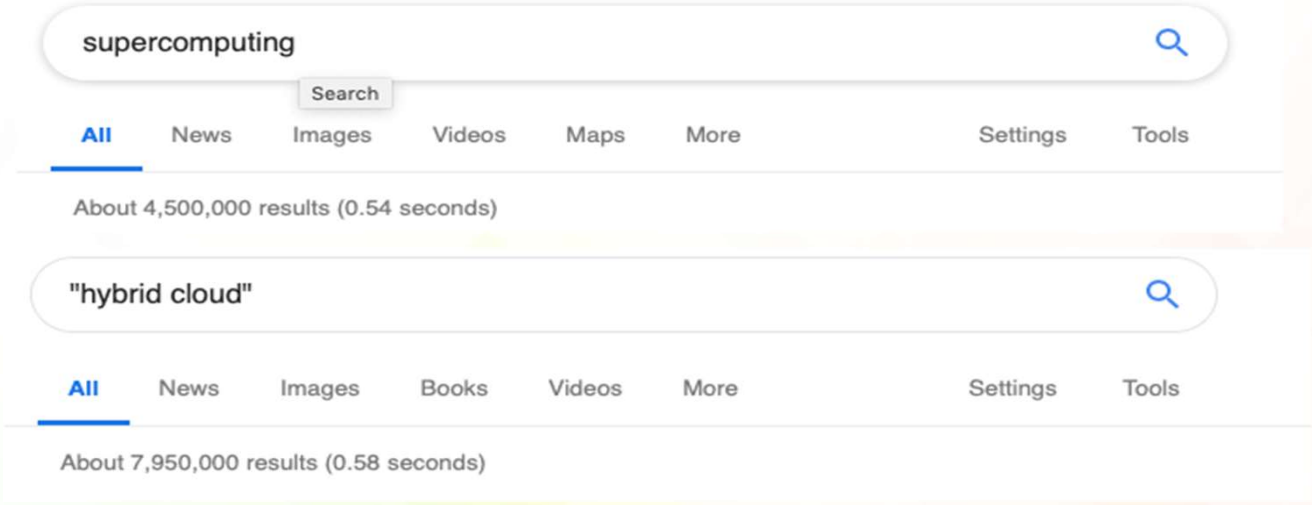
It is built upon modern SDN and cloud computing technologies, and provides secure and high-performance services in a multi-tenancy, on-demand and as-a-service fashion.

Software Defined Networking @MAX

- DRAGON Research Network (production)
- GENI Network and GENI Rack (production)
- Software-Defined ScienceDMZ (protoduction)
- Software-Defined WIX Exchange (prototype)
- OpenFlow and SDN Testbeds (research)
- SENSE Testbed (research)

Cloud Economics





A Need for “Edge”

- Edge Computing: A distributed computing paradigm focused on processing data at local compute servers and devices, instead of transmitting them to data centers.
- Edge Cloud: A distributed cloud facility (or scaled-down data center) at network edge.
- ScienceDMZ: A widely adopted “edge” networking solution for big science data movement for the research community.
- For both economic and performance reasons, our community needs a cloud-like ScienceDMZ facility that adopts edge computing paradigm and is customized for scientific research workflows.

A Chronicle of Ideas

- MAX was among the first to use AWS Direct Connect service at its eastern region.
- In 2013, we started experimenting provisioning multi-tenancy services over a 10G AWS Direct Connect pipe, dynamically.
- An extension of the DRAGON control plane, plus BGP and VRF tricks – a Cloud Connect precursor.

SDN, Cloud and Edge

- The NSF SDNX project (2012-2014) promoted the idea of **bringing applications to network** edge.
- The NSF HPCDNA project (2013-2015) conceived the idea of **network-embedded compute and storage** and started integrating SDMZ, HPC and AWS over SDN, to break into the "walled garden".

Infrastructure Datafication

- The DOE RAINS project (2014-2017) developed the Multi-Resource Markup Language (MRML) and laid the foundation for **model driven full-stack automation and orchestration**.
- The NSF RECAST project (2015-2017) leveraged RAINS technologies and developed the **Software Defined ScienceDMZ (SD-SDMZ)** architecture.
- SD-SDMZ = cloud-like, regionalized / shared, as-a-service model.

Big Orchestration

- The DOE SENSE project (2016-2019) adopted RAINS modeling and computation tools, and developed the SENSE software stack for **distributed and scalable end-to-end orchestration**.
- MAXedge is on the converging paths of SD-SDMZ architecture and SENSE orchestration software.
- In 2018, MAX rolled out the **Advanced Hybrid Cloud** (AHC) and **Edge Cloud Connect** (ECC) pilot services.

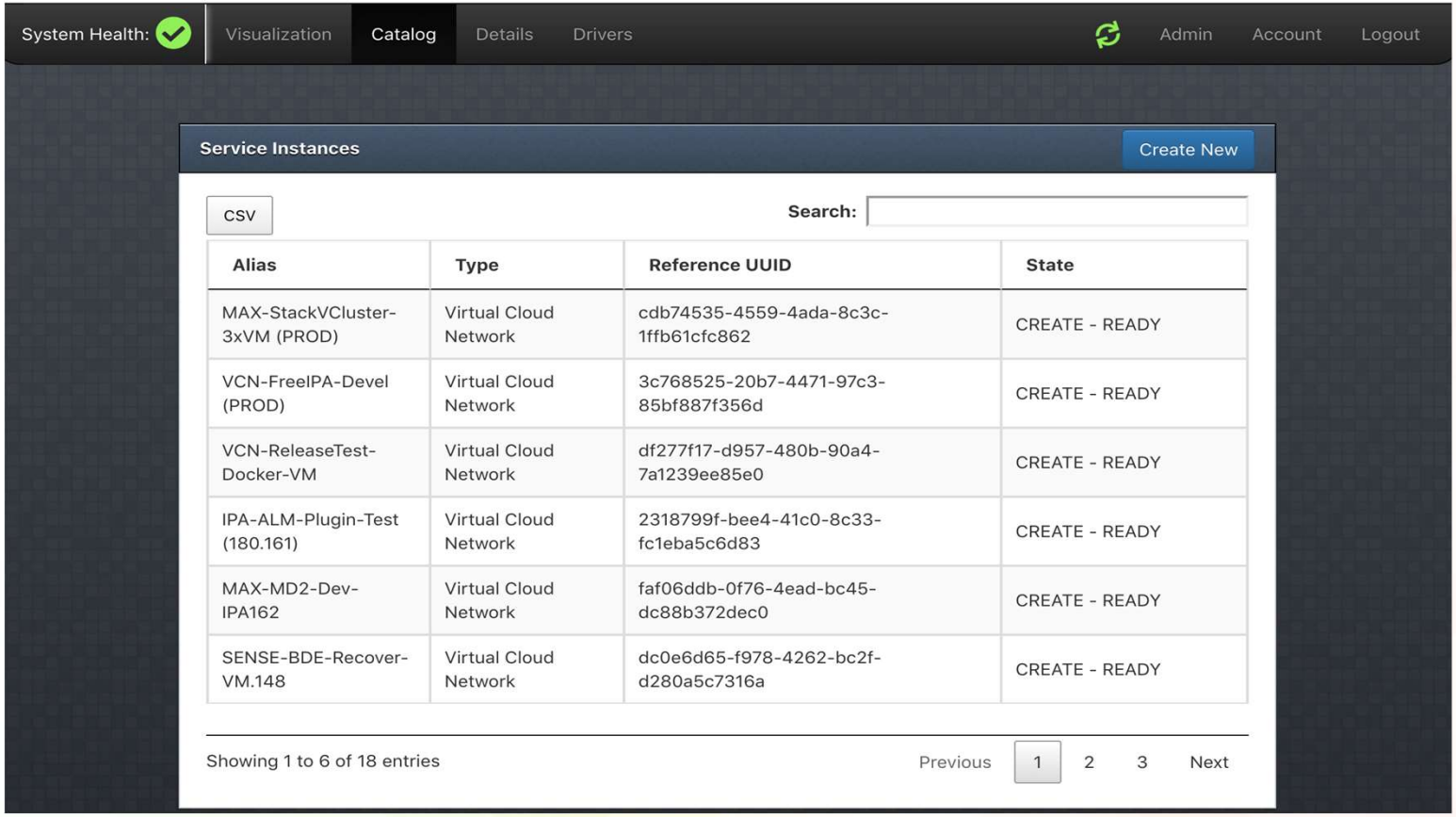
The MAXedge

- An SD-SDMZ built on SENSE software stack.
- A model-driven end-to-end service orchestrator.
- Arbitrary combination of SDN, Cloud, HPC and Edge compute and storage resources based on abstract user intents.
- Secure, high-performance, dedicated and API-driven services for end-to-end workflows.

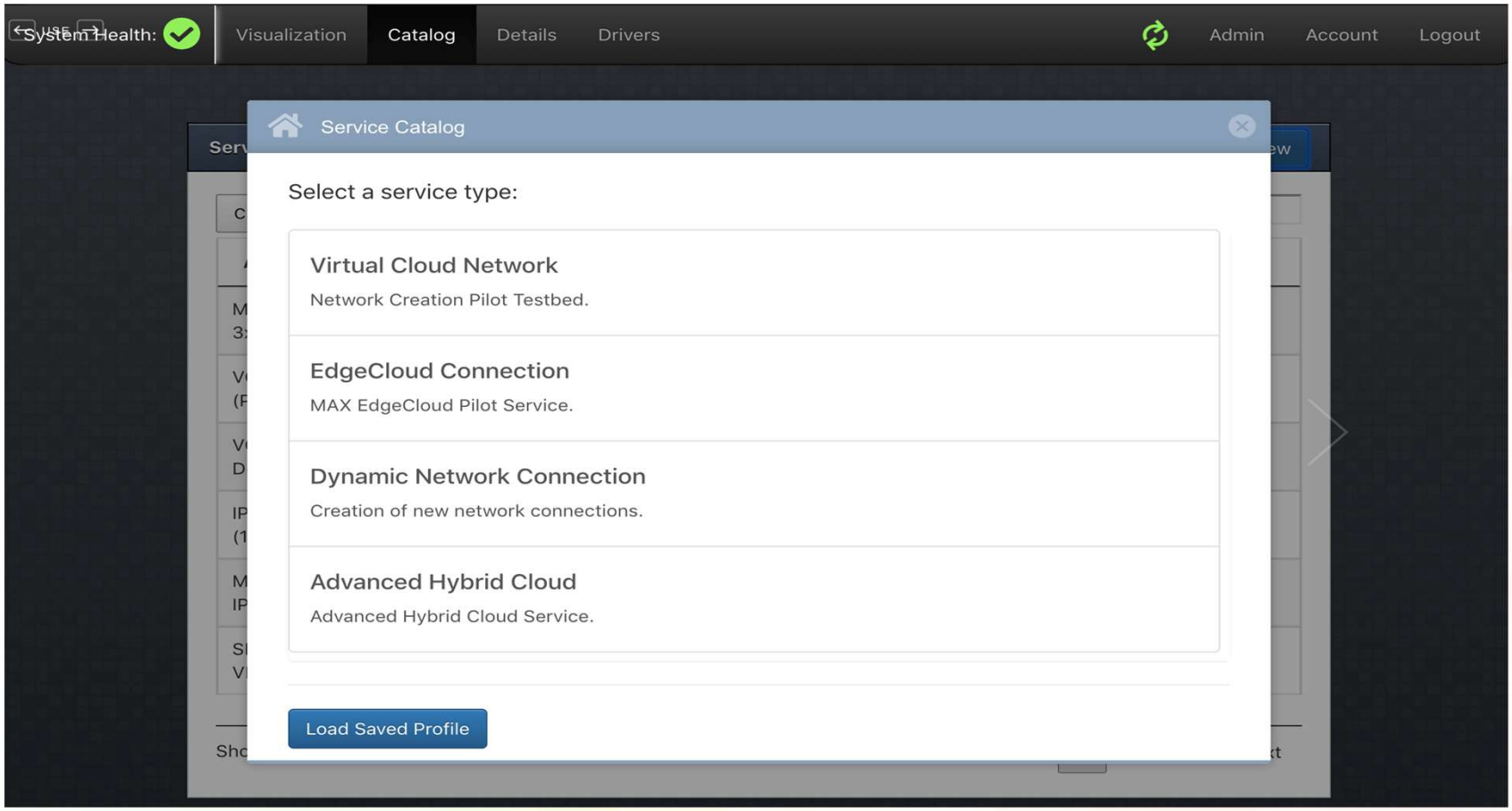
Outline

- What is MAXedge?
- How do we use MAXedge?
- What services does MAXedge offer?
- How does MAXedge work? – A deep dive

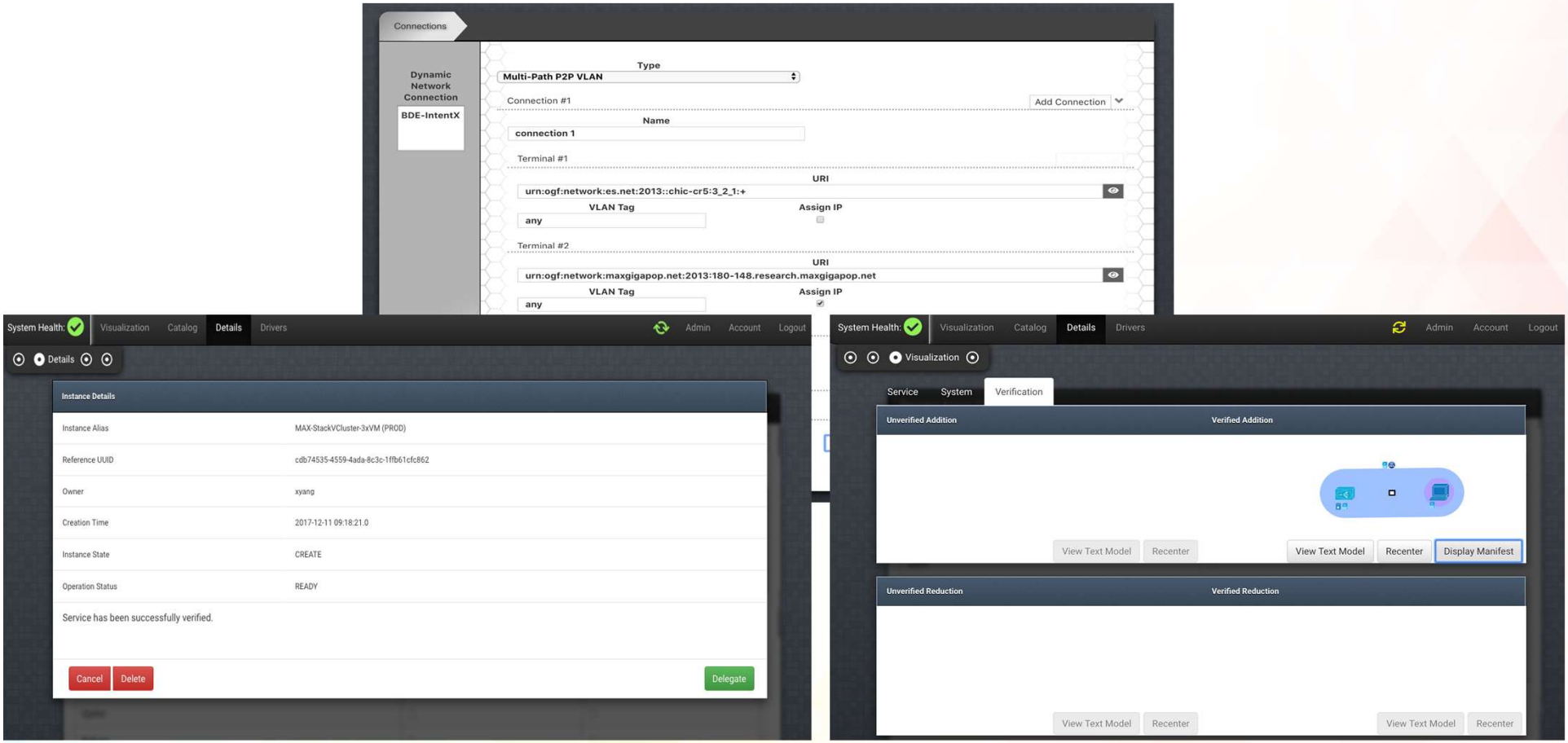
Orchestrator Portal



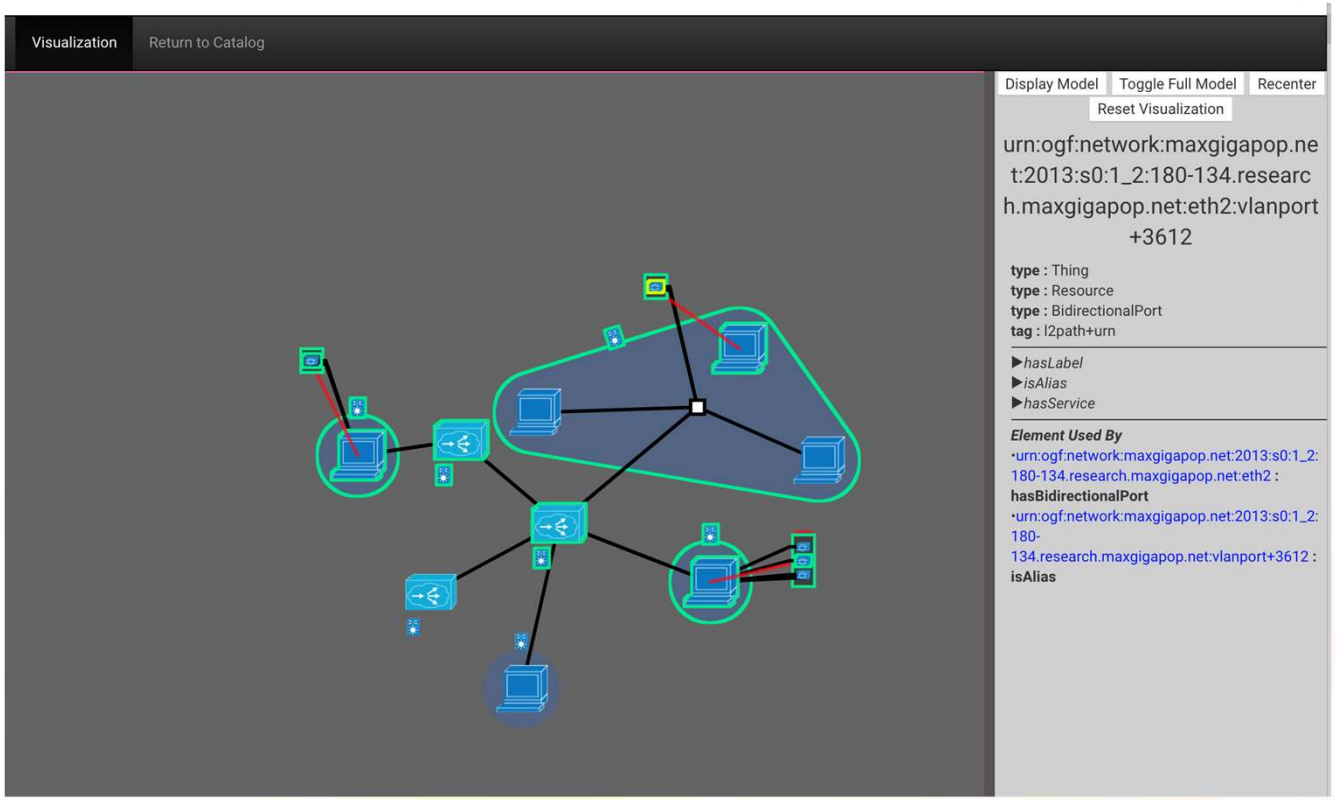
Service Catalog



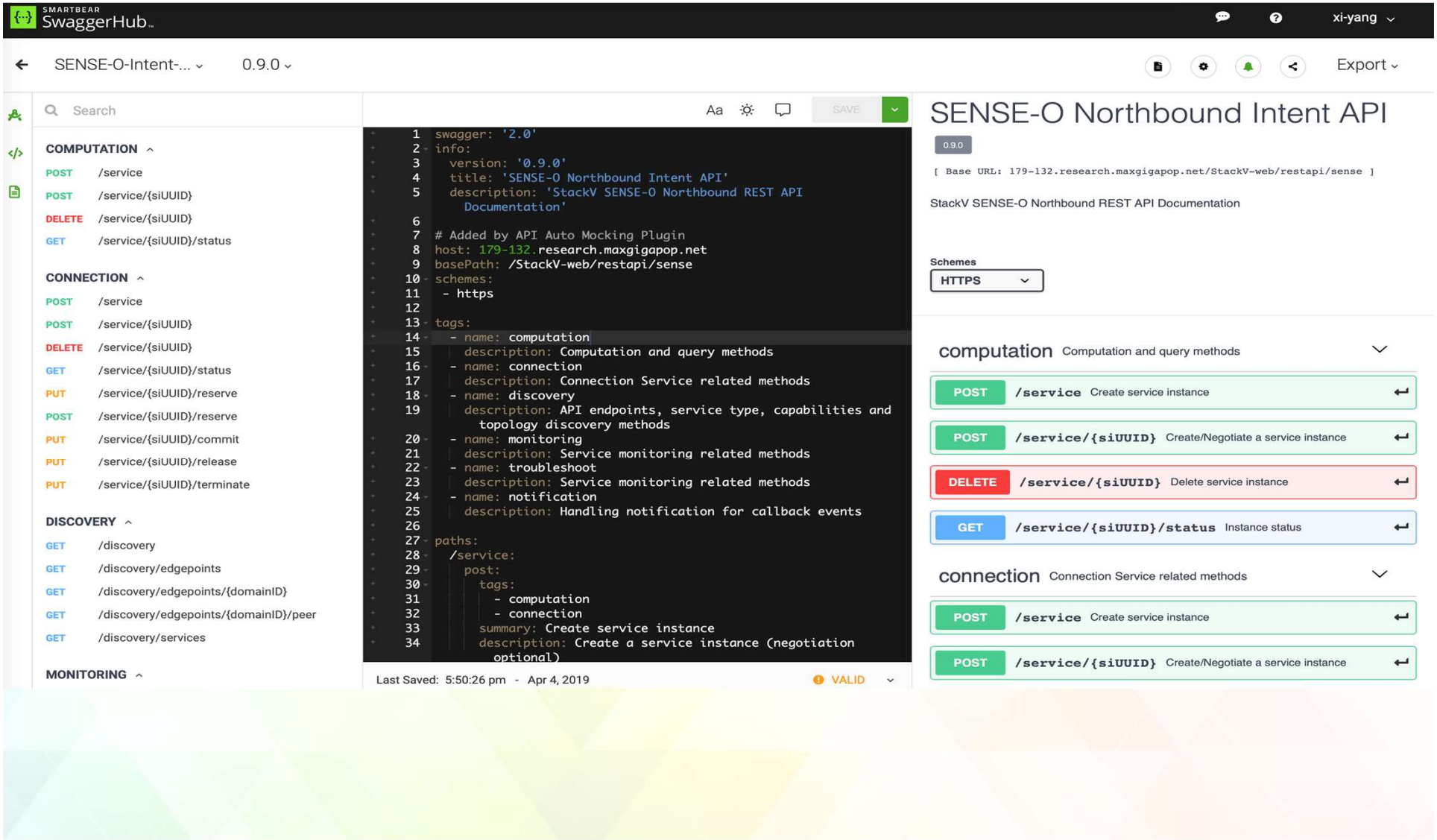
Service Intent and Lifecycle Management



Infrastructure and Service Visualization



API Driven Interaction for App Workflows



Outline

- What is MAXedge?
- How do we use MAXedge?
- What services does MAXedge offer?
- How does MAXedge work? – A deep dive

Hybrid Cloud for HPC + Big Data?

- Cloud services are doing great for Big Data,
  - Applications such as Hadoop at large scale
- But they still cannot deliver the performance many high-end apps need.
  - limited type of resources by public providers
  - performance capped by virtualization and inter-cloud networking
  - It just does not feel like your high performance computing, parallel storage, high speed data transfer and deterministic speeds at “home”.

Advanced Hybrid Cloud (AHC)

Cloud-Like Service + HPC Experience

- ✓ **Orchestrated hybrid cloud with deterministic end-to-end high performance**
- ✓ **Local OpenStack Virtual Machines (with SRIOV interfaces to network and storage)**
- ✓ **Dedicated Local Ceph Storage Resources and Connections (SRIOV interface)**
- ✓ **Integrated AWS Resources (VPC or Public over Direct Connect)**
- ✓ **Dedicated Network Connections with QoS**
- ✓ **Set up in one click "as-a-service"**

User Supplies the following information

1) Local Cloud Resources

- **number of VMs**
- **amount of storage**

2) Public Cloud (AWS) Resources

- **Direct Connect access**
  - **Shared (with preemption)**
  - **Dedicated (bandwidth, schedule (start time, end time))**
- **AWS account number (User or MAX?)**
  - **If User AWS account number that is all that is needed**
  - **If MAX AWS account number, specify AWS Resources, VPC, Instance Type, Storage amount, S3 Endpoint**

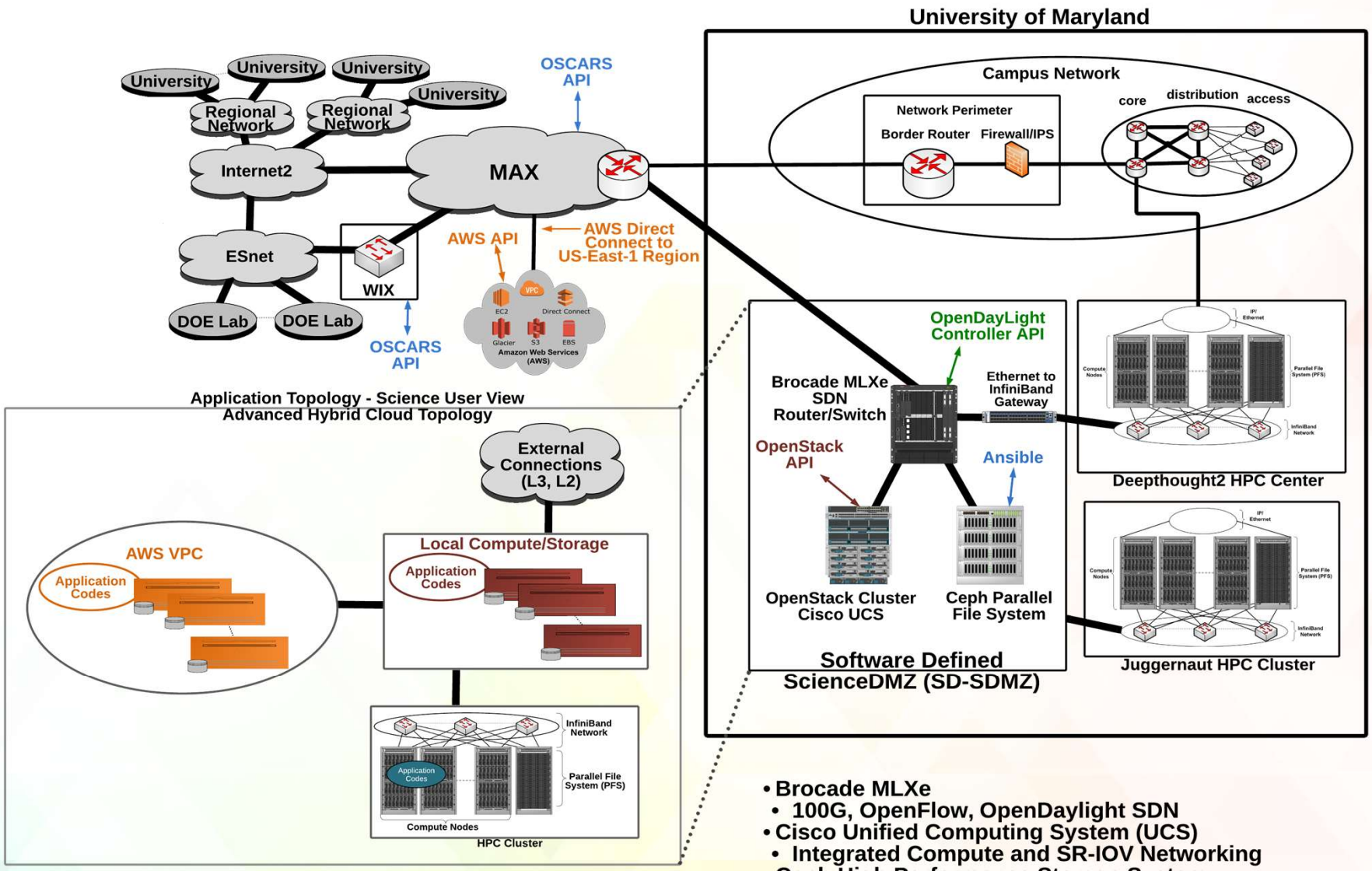
3) External Connections (urn list discoverable from MAXedge portal or API)

- **AL2S (remote endpoint)**
- **Other external resources as available**

4) Service Level Parameters

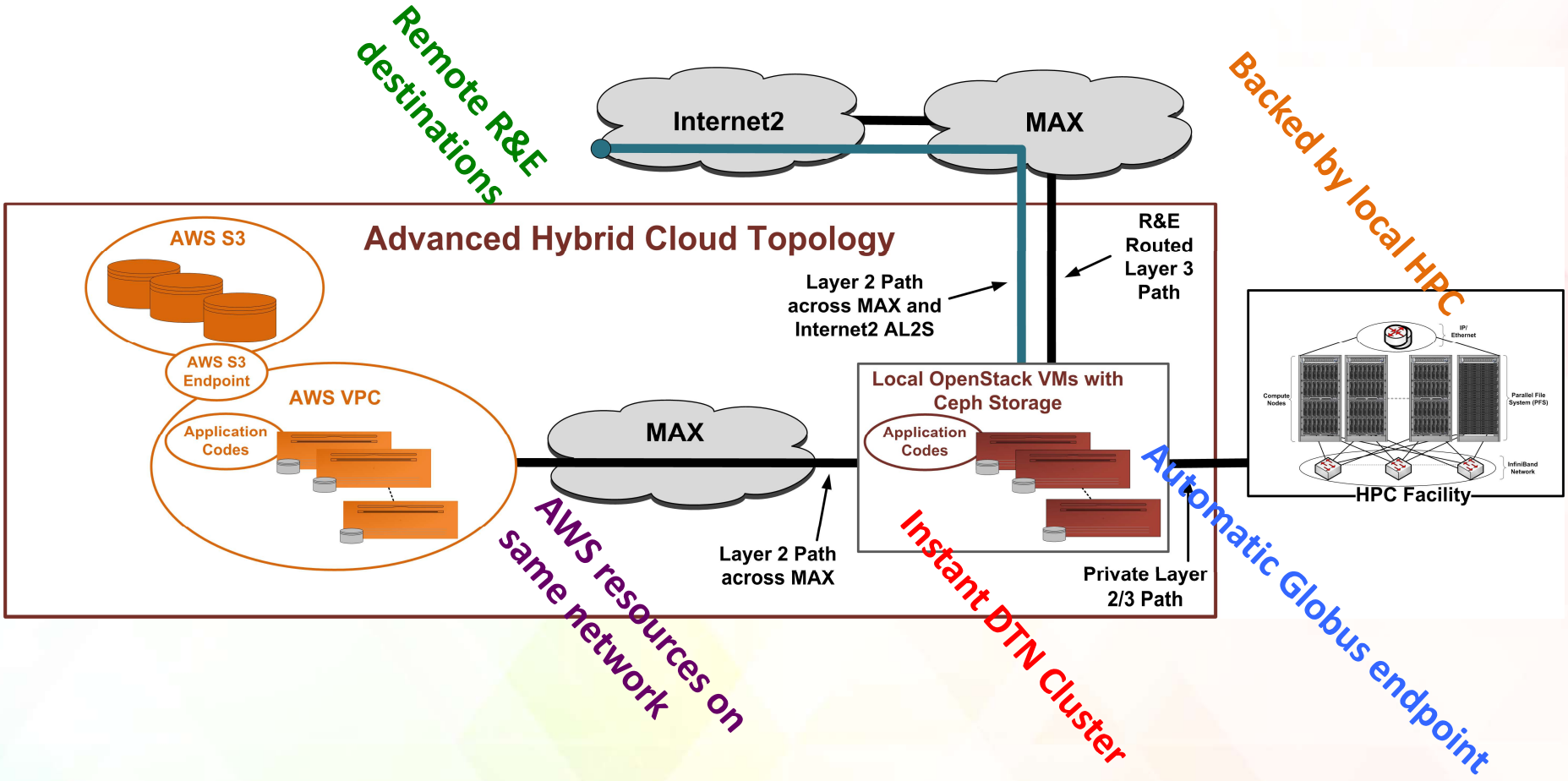
- **Define number of simultaneous service instances allowed**

MAXedge AHC-Ready Infrastructure



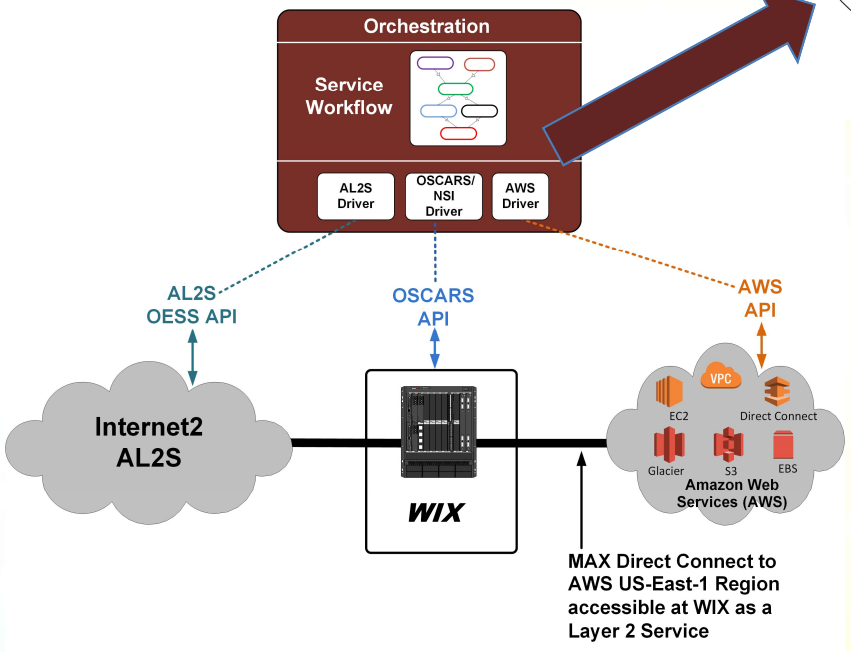
- Brocade MLXe
  - 100G, OpenFlow, OpenDaylight SDN
- Cisco Unified Computing System (UCS)
  - Integrated Compute and SR-IOV Networking
- Ceph High Performance Storage System

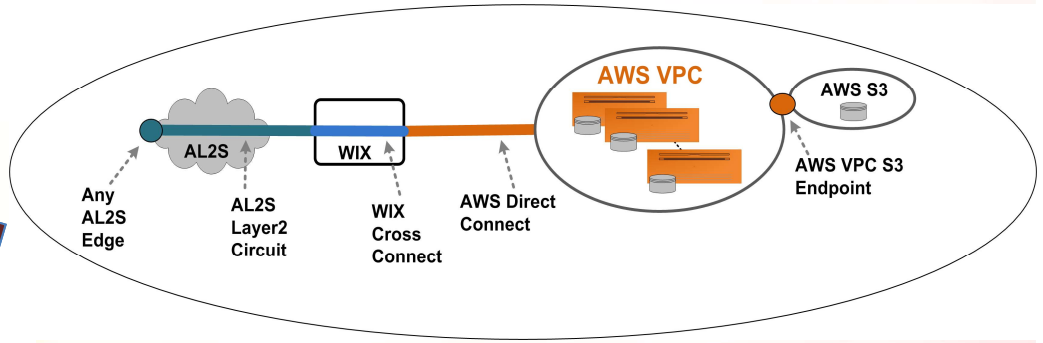
MAXedge AHC Service in One Click



Edge Cloud Connect (ECC) Service

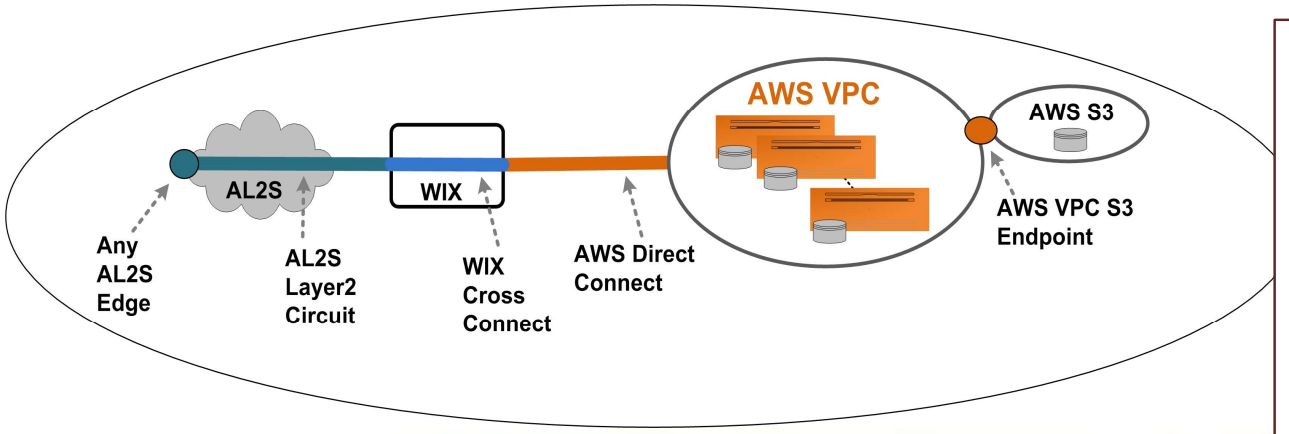
MAXedge orchestrates AL2S, WIX, and AWS Direct Connect provisioning





Dedicated layer-2 connection with QoS

Edge Cloud Connect (ECC) Service



- Available: AWS Direct Connect via shared or scheduled/dedicated mode
- Not available: compute, storage, BGP instance for AWS peering

User supplies the following information

1) Public Cloud (AWS) Resources

- **Direct Connect access**
  - **Shared (with preemption)**
  - **Dedicated (bandwidth, schedule (start time, end time))**
- **User AWS account number**

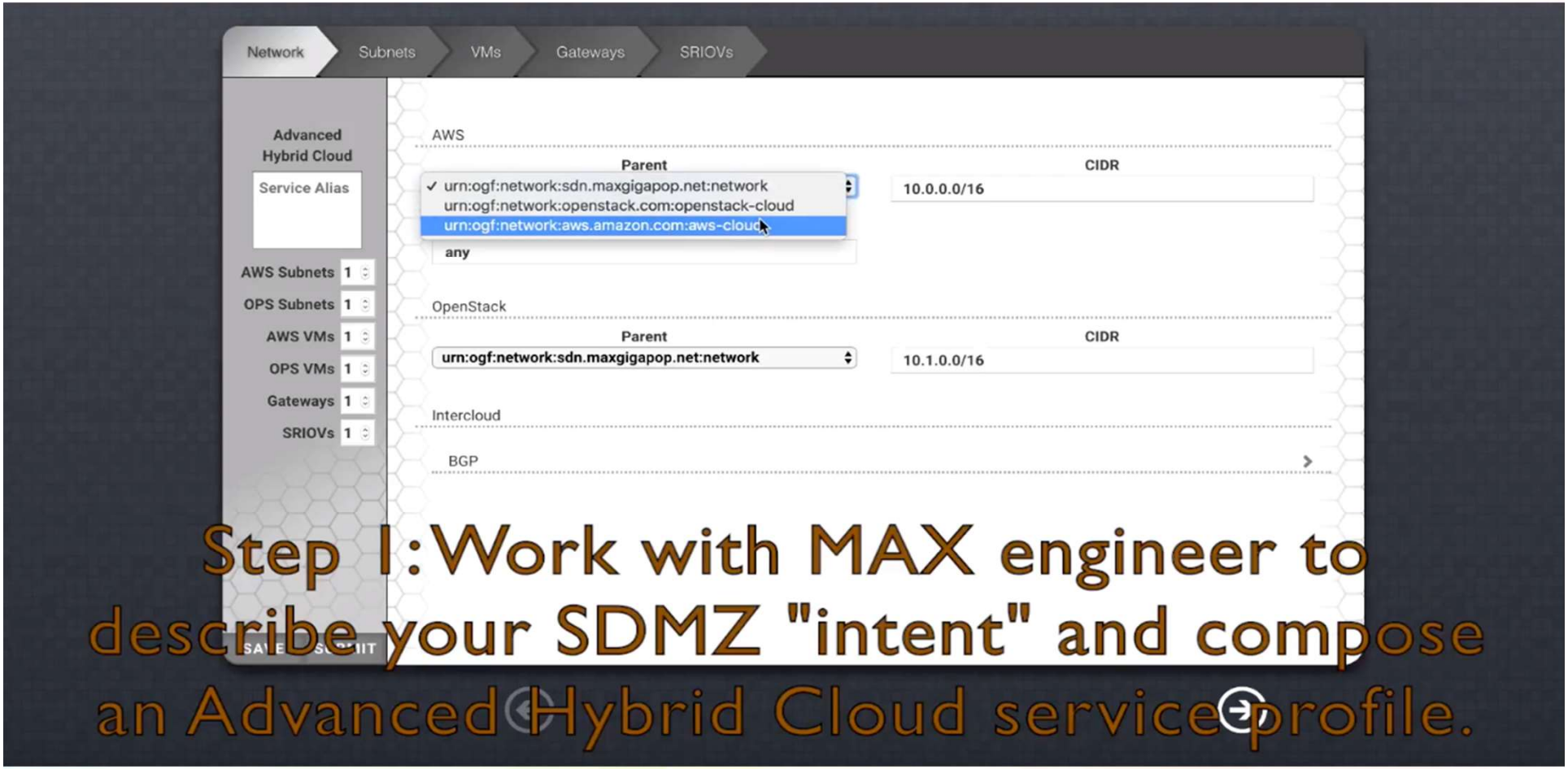
2) External Connections

- **AL2S (remote endpoint)**

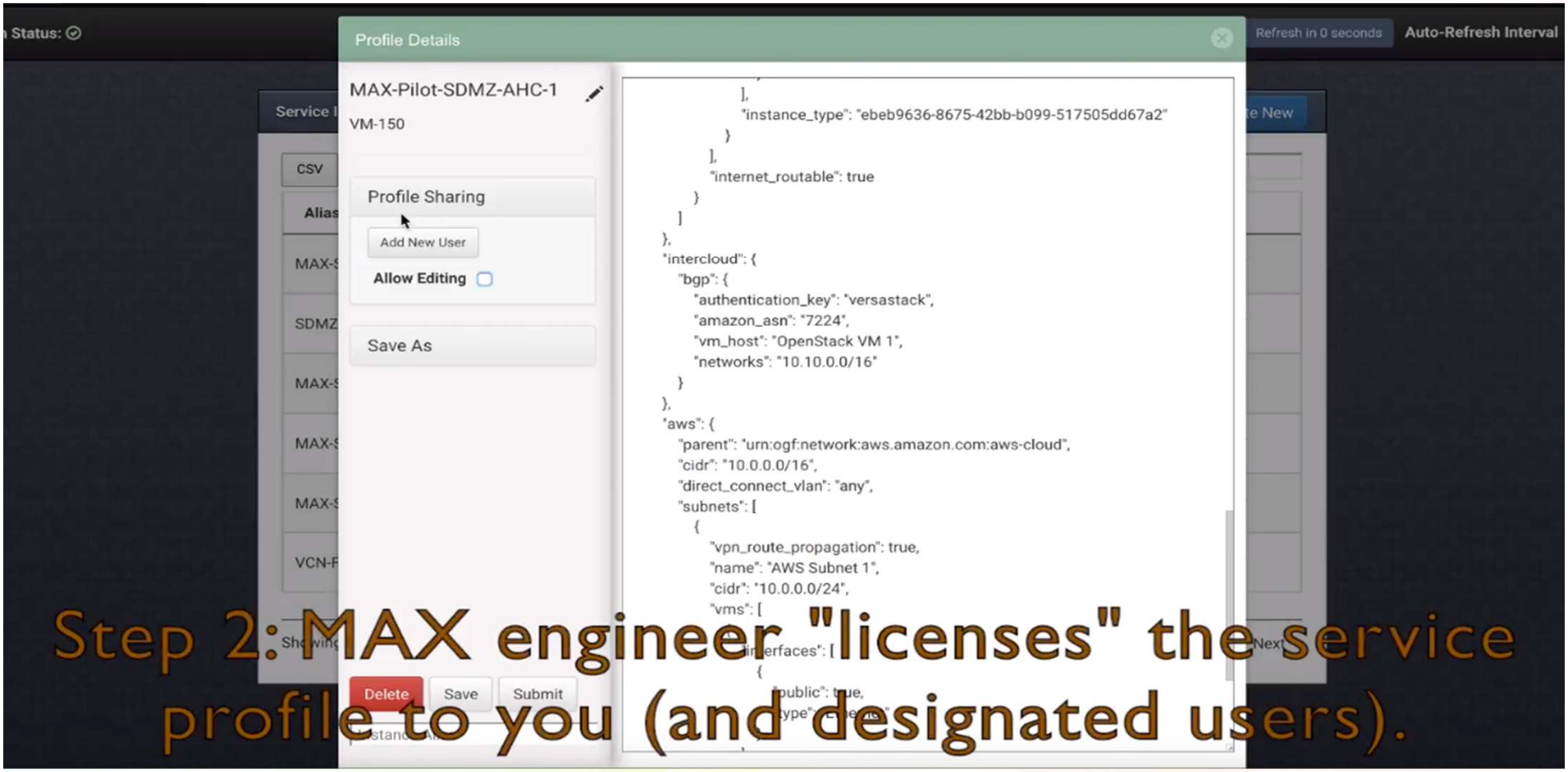
3) Service Level Parameters

- **Define number of simultaneous service instances allowed**

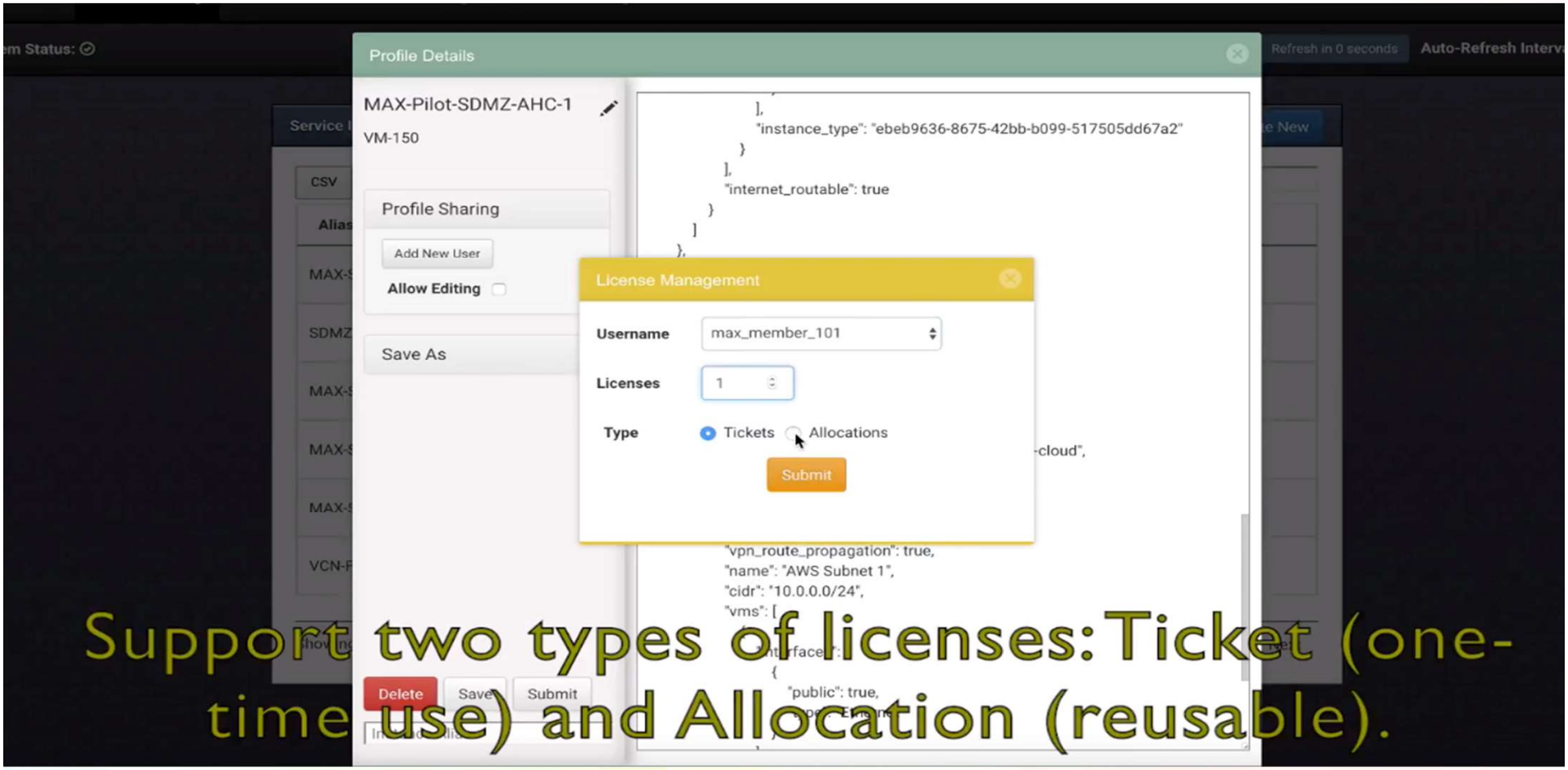
MAXedge Service Workflow for Customers



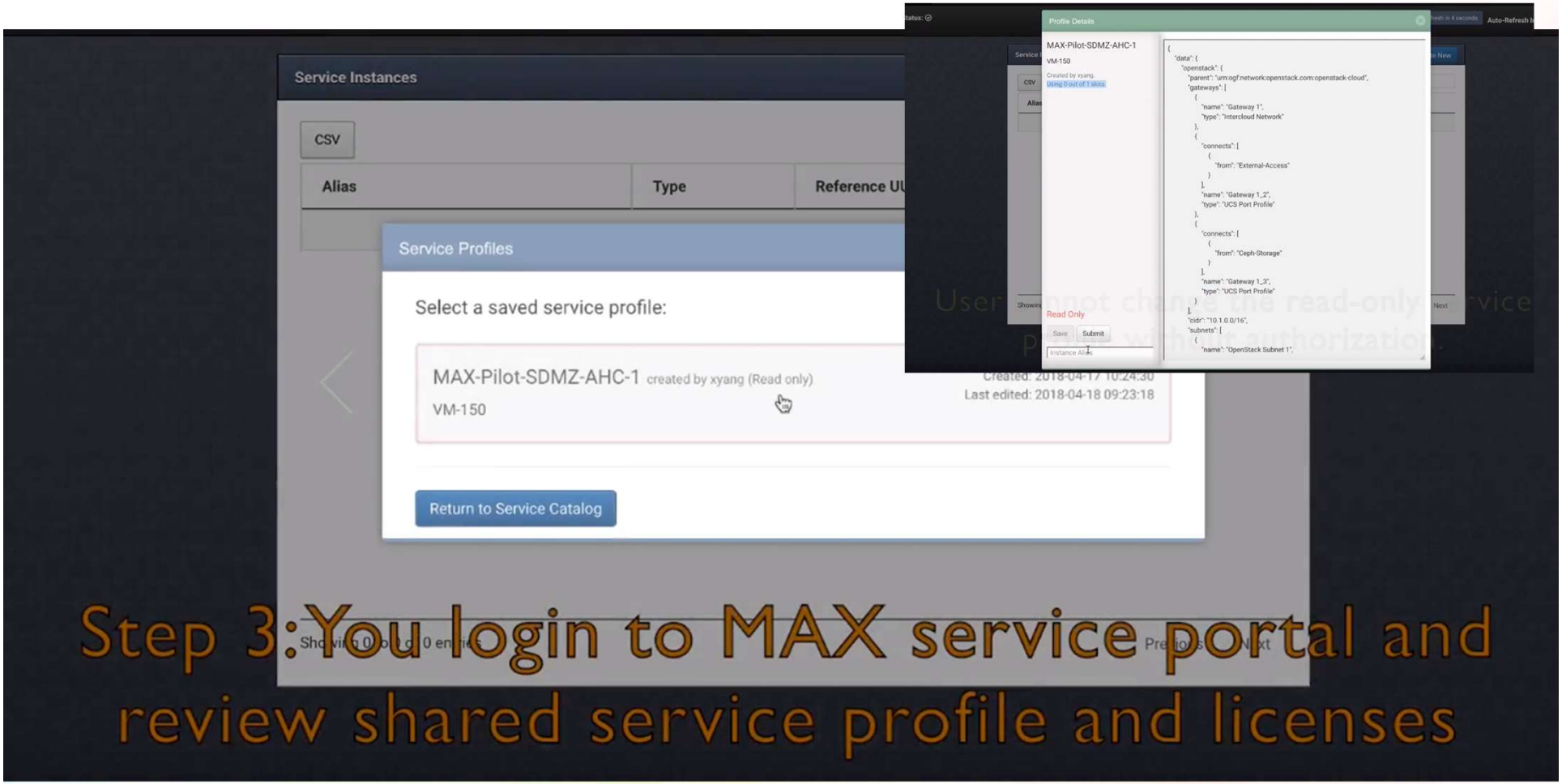
MAXedge Service Workflow for Customers



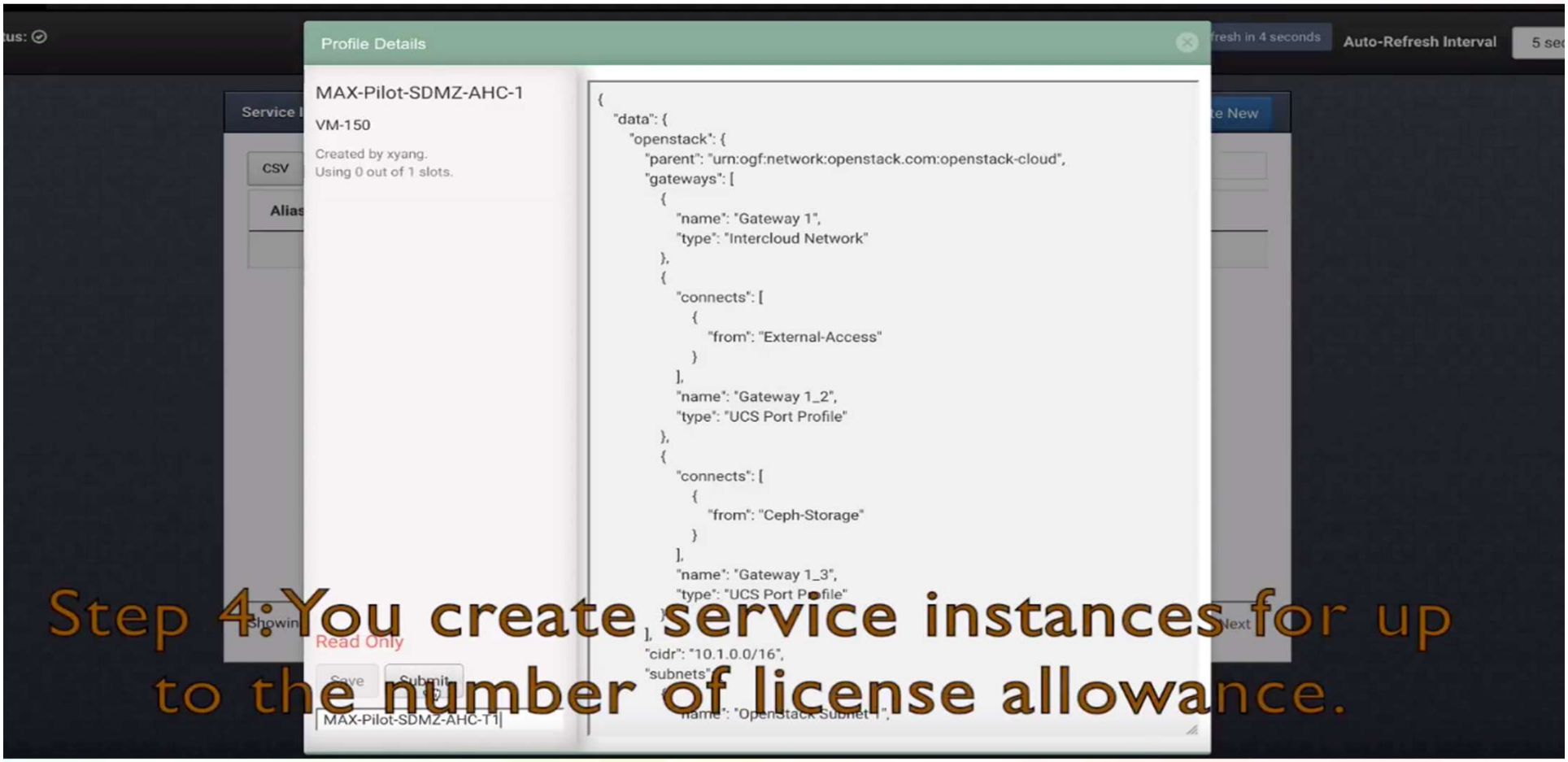
MAXedge Service Workflow for Customers



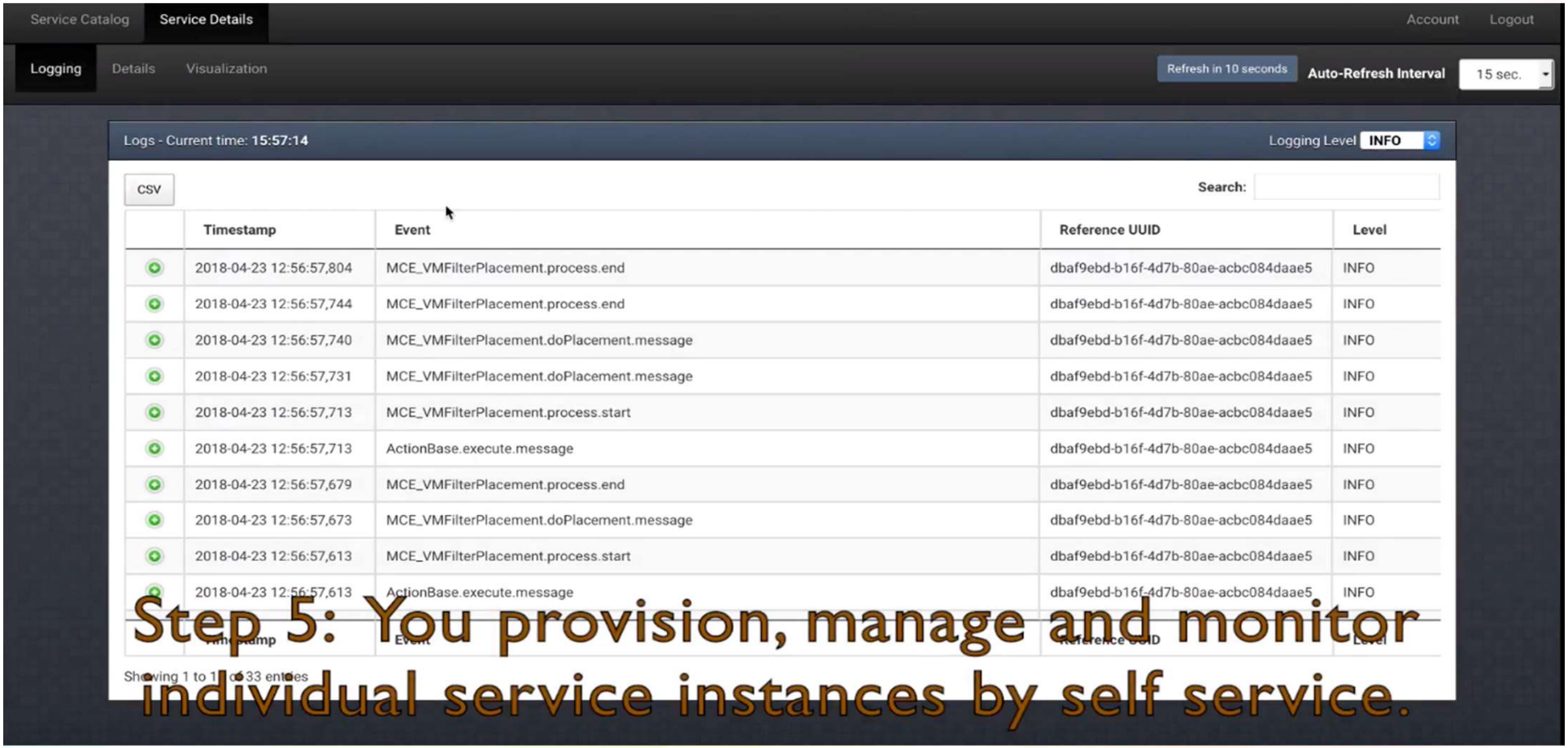
MAXedge Service Workflow for Customers



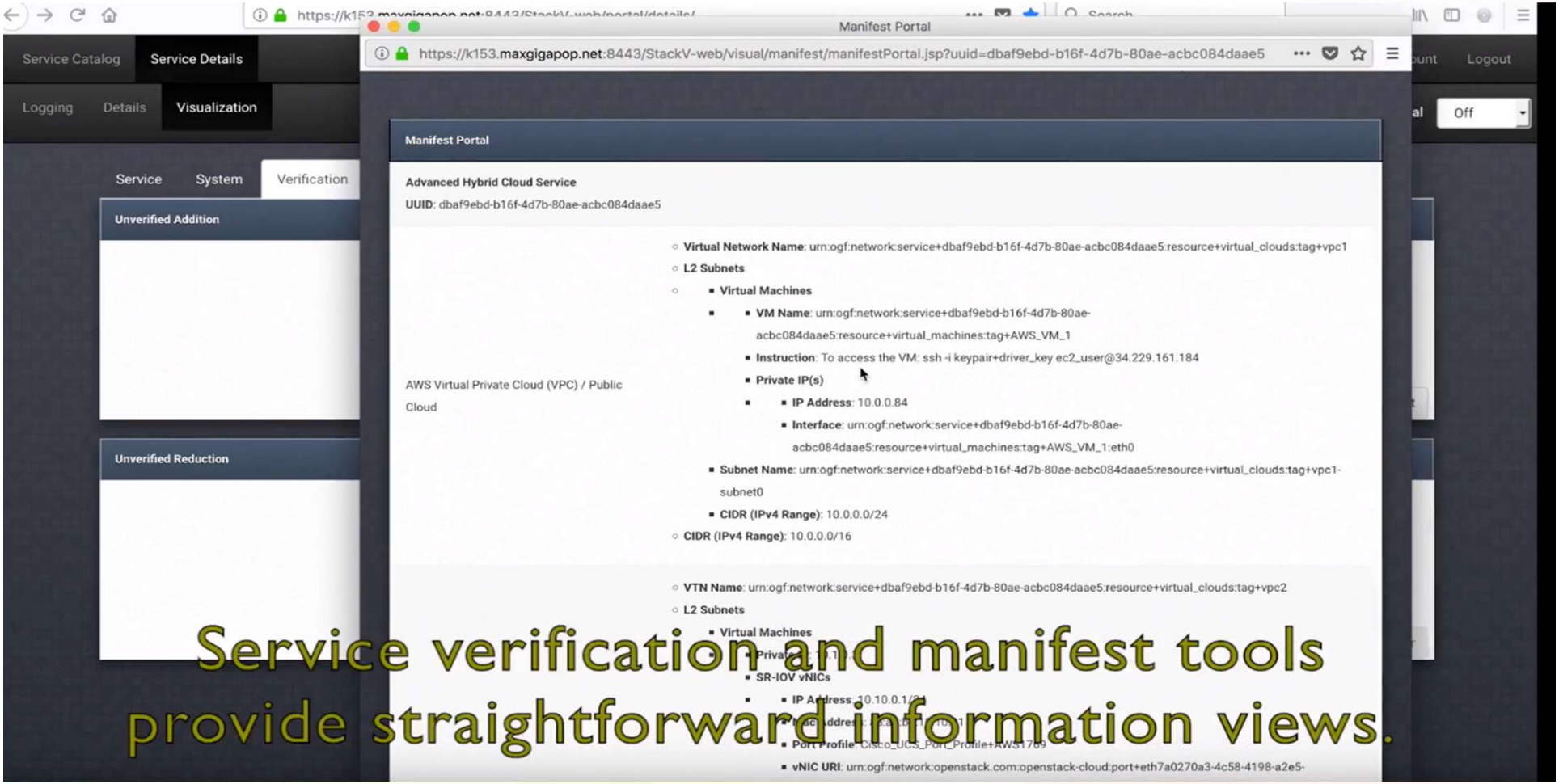
MAXedge Service Workflow for Customers



MAXedge Service Workflow for Customers



MAXedge Service Workflow for Customers

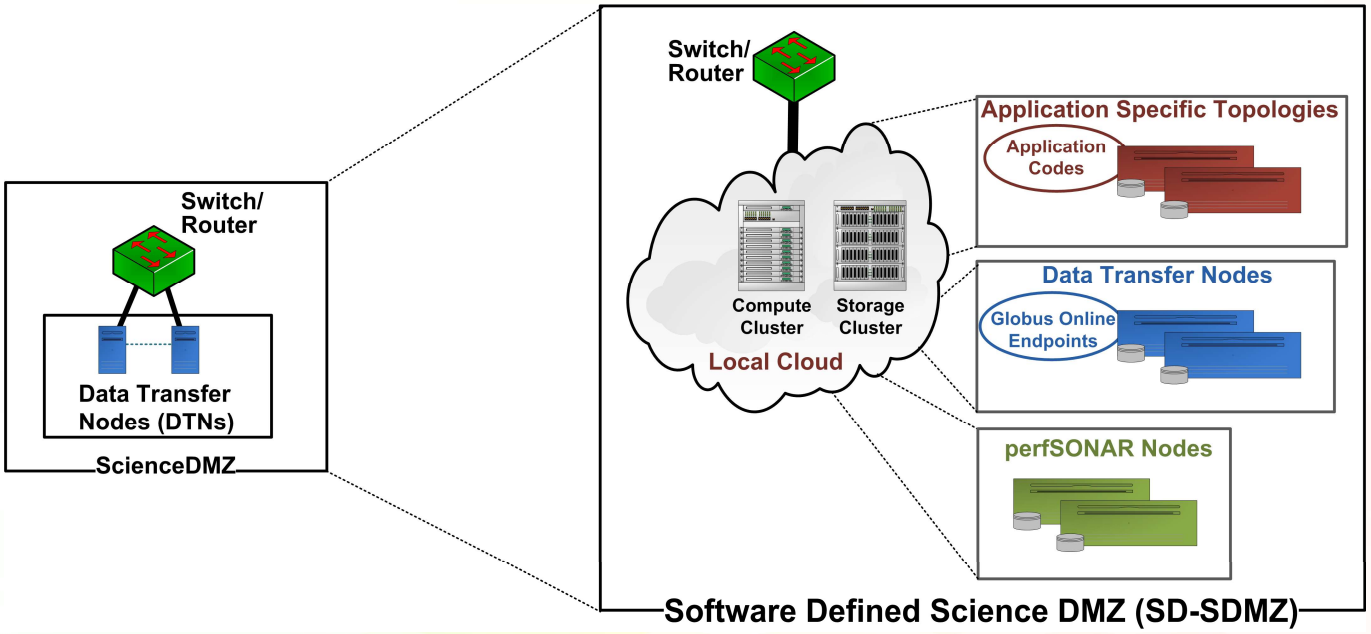


Outline

- What is MAXedge?
- How do we use MAXedge?
- What services does MAXedge offer?
- How does MAXedge work? – A deep dive

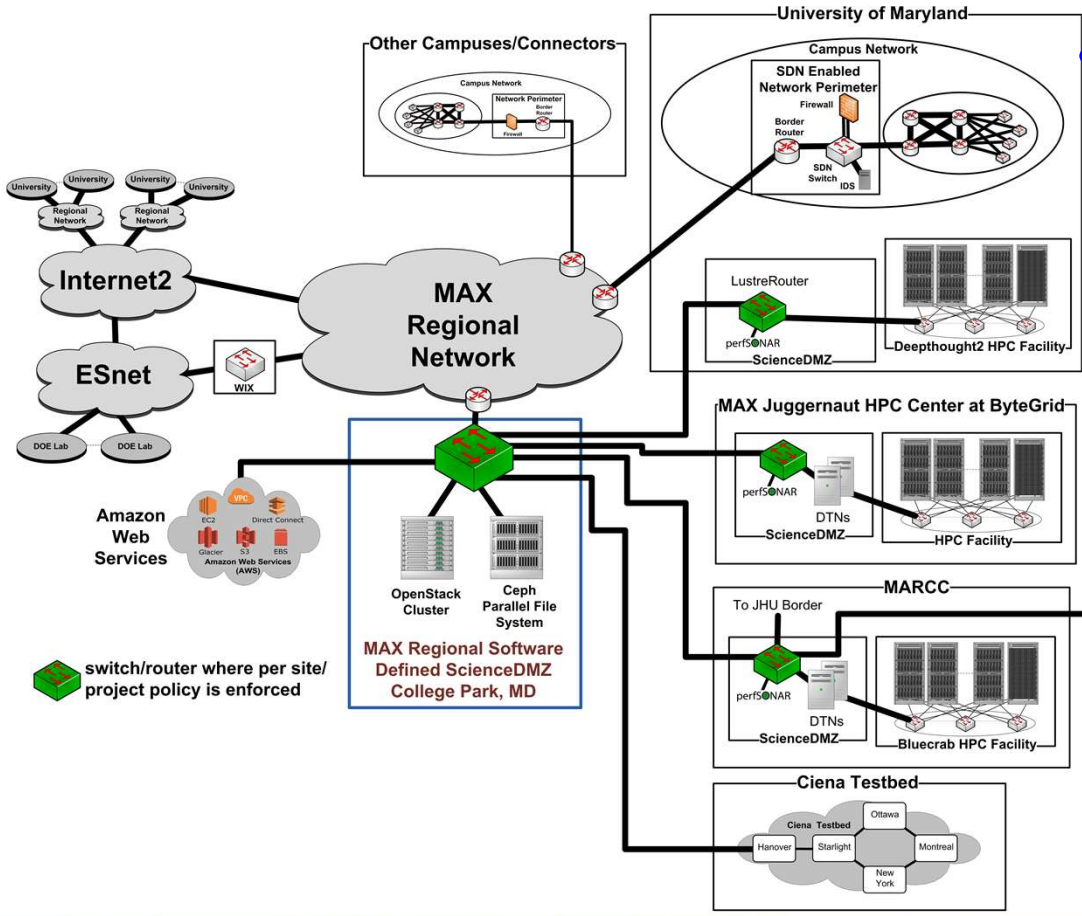
MAXedge: Virtualize and Cloudify

- ScienceDMZ transitions to local cloud compute, storage, SDN control. On-demand, scalable Hybrid Cloud services.



- Bare metal to local cloud evolution enabled by full stack automation and orchestration

MAXedge: Regionalize and as-A-Service



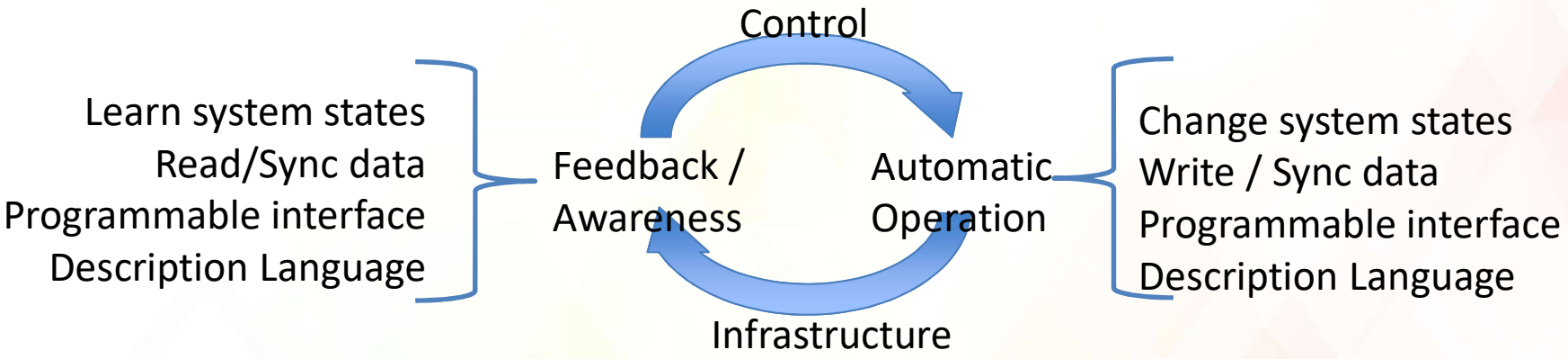
- Leverage MAX high performance networks to service the regional community
- Transition to “as A Service” model using “web services” to support “web scale” operations

Technical Approach

- Multi-Resource Orchestration
  - integrating and orchestrating the network and network services with the things that attach to the network – compute, storage, clouds, and instruments.
- Full-Stack Model Driven
  - using semantic models to describe resources in order to allow integrated reasoning, abstraction, and user intents, consistently through all layers of the system stack.
- Intelligent Computation Services
  - pluggable Model Computation Modules and “composable intelligence” to orchestrate “arbitrary” services in response to high level user requests.
- “Orchestrate the Automaters”
  - Lots of automations work together, end-to-end

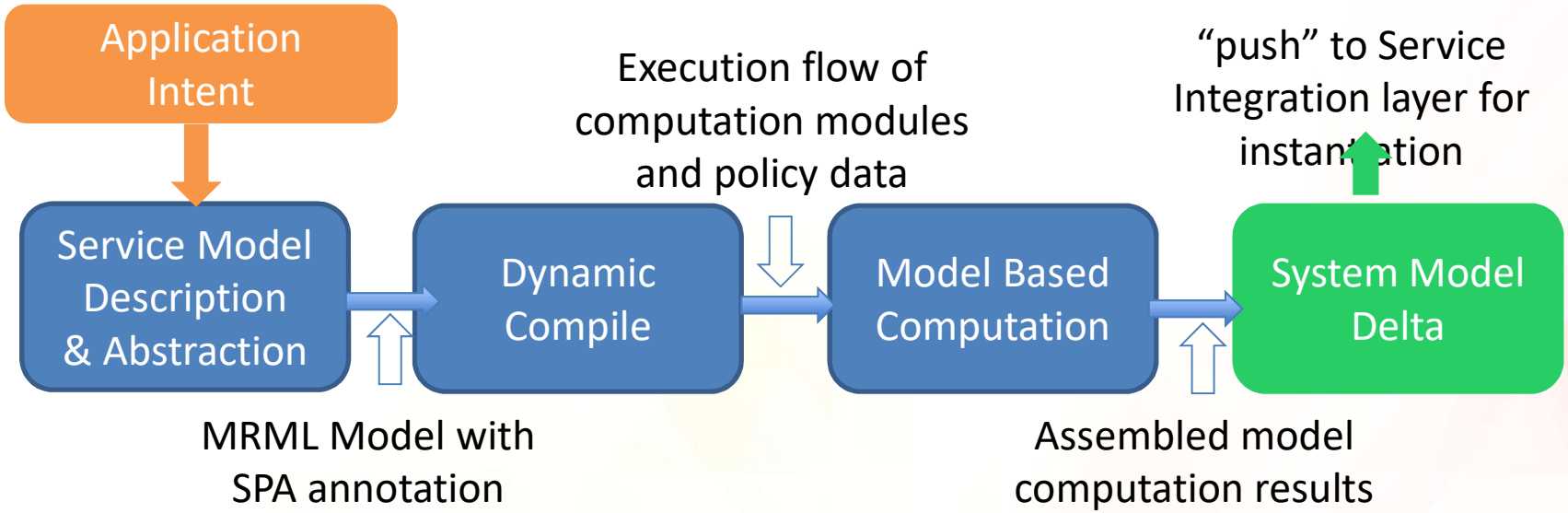
Model Based Orchestration

It is all about Automation and Orchestration in a Control-Feedback loop.

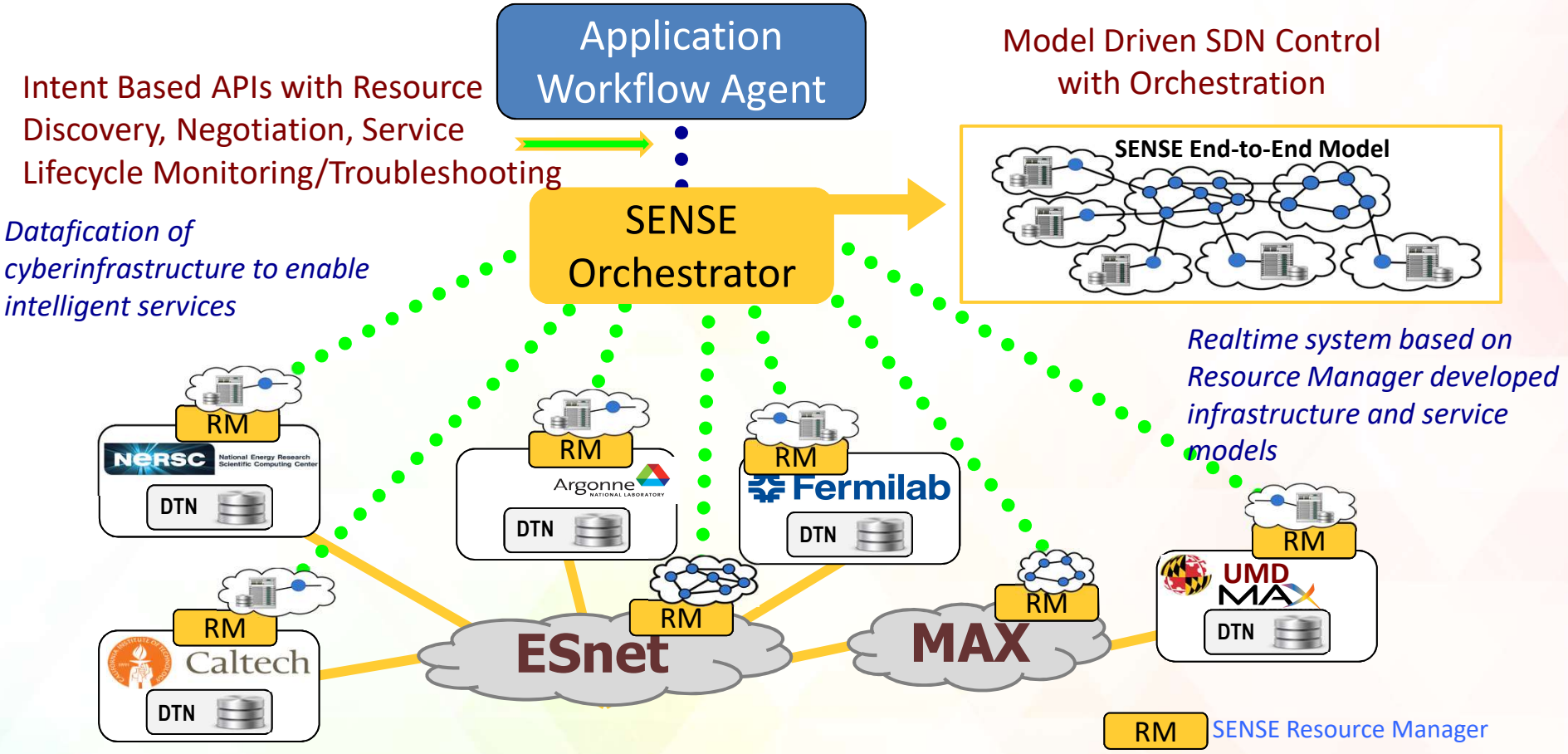


Objective is to allow the machines to automate, iterate, react, and adjust to find solutions and not bring the humans in until absolutely necessary

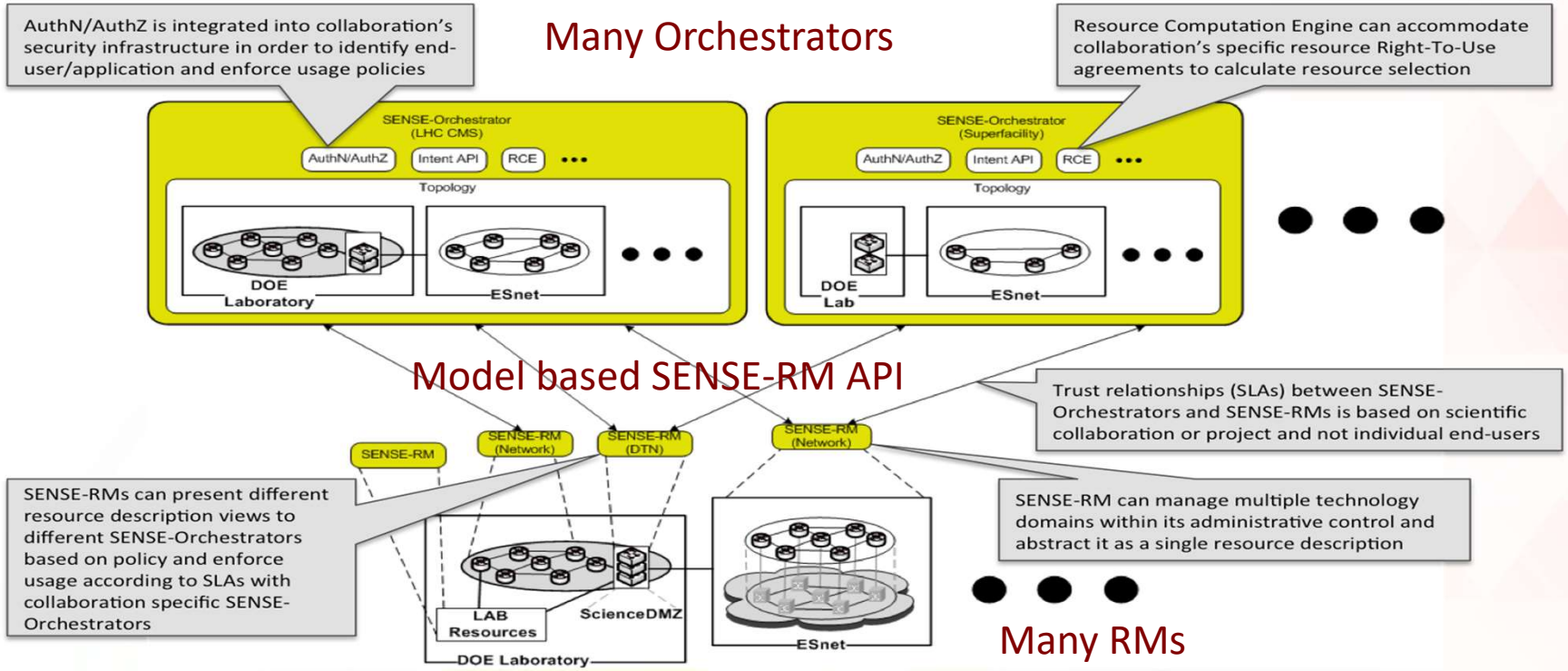
Service Orchestration Workflow



MAXedge Is Built on SENSE Architecture

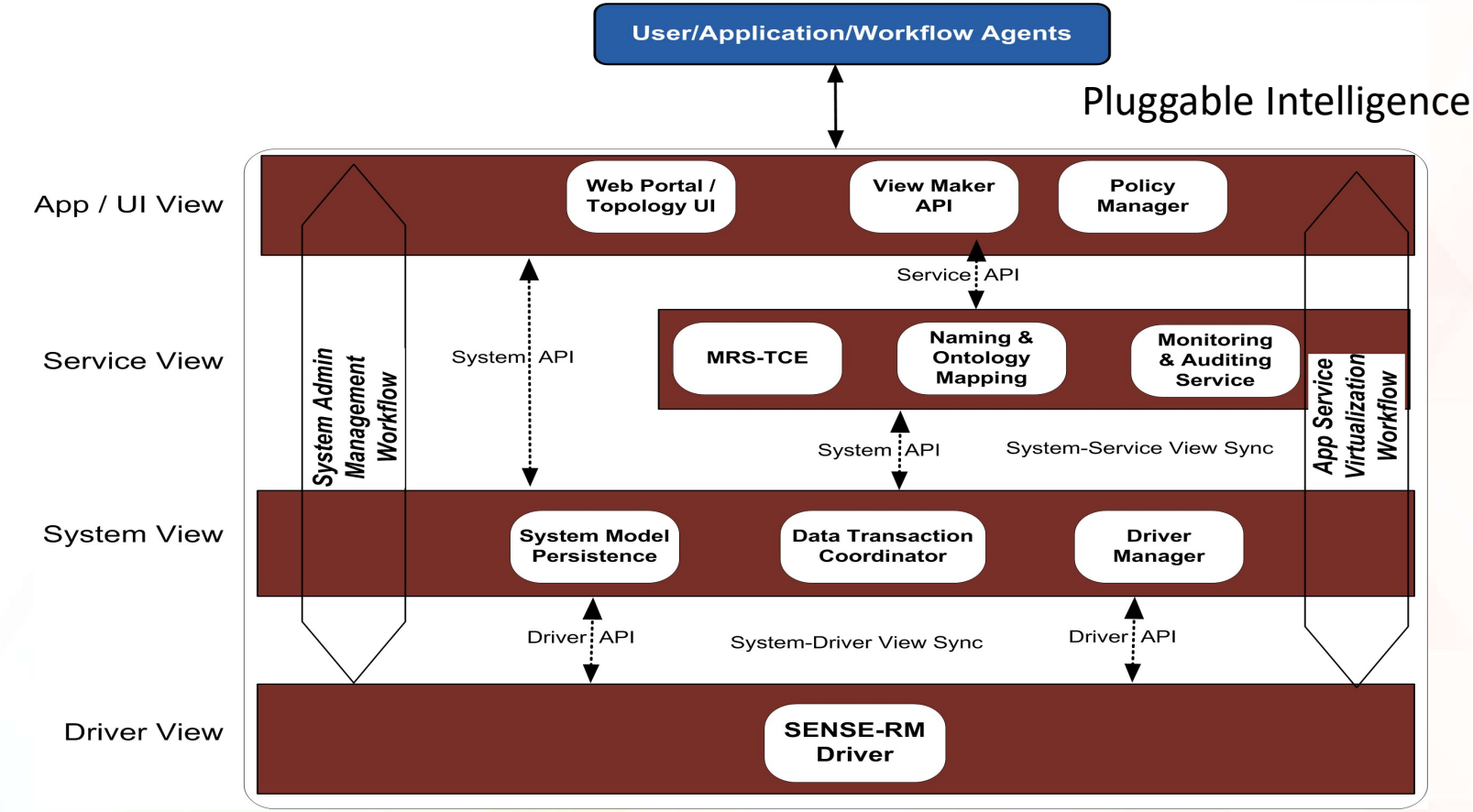


Many-to-Many Service Oriented Architecture

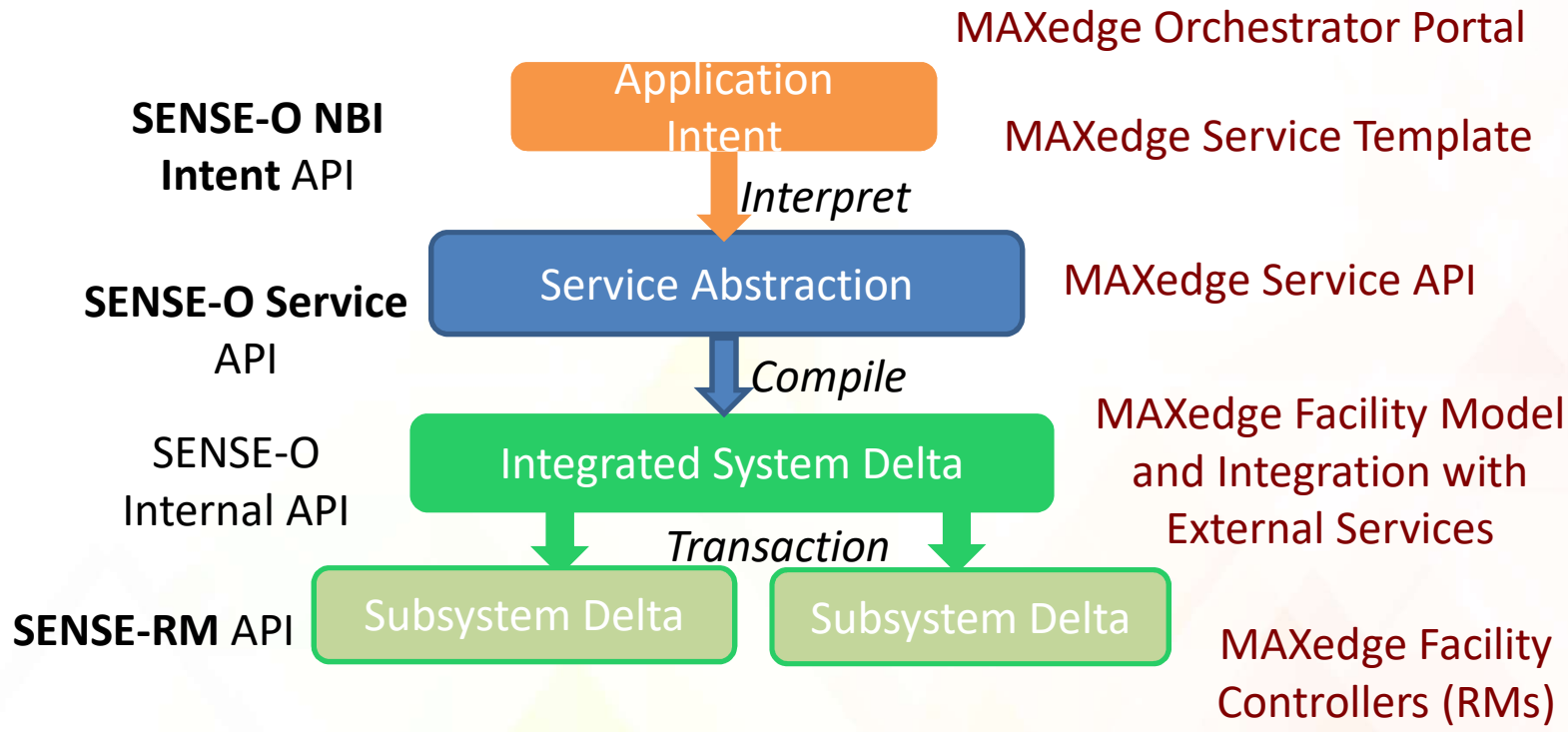


SENSE Orchestrator Software

SENSE Open Source Software Stack



MAXedge: A Local SENSE System Instantiation



Modernize an SDMZ / Facility / Site with the SENSE stack …

MAXedge and MAX Research

- UMD/MAX R&D Team
  - Xi Yang
  - Alberto Jimenez
  - Multiple Students
- Results from several research projects including:
  - SDN for End-to-End Networked Science at the Exascale (SENSE)
  - Regional Embedded Cloud for As-a-Service Transformation (RECAST)
  - Resource Aware Intelligent Network Services(RAINS)
  - High Performance Computing with Data and Networking Acceleration (HPCDNA)
  - Software Defined Network Exchange (SDNX)
  - GENI Enabled Software Defined Exchange (SDX)

U.S. DEPARTMENT OF ENERGY
Office of Science

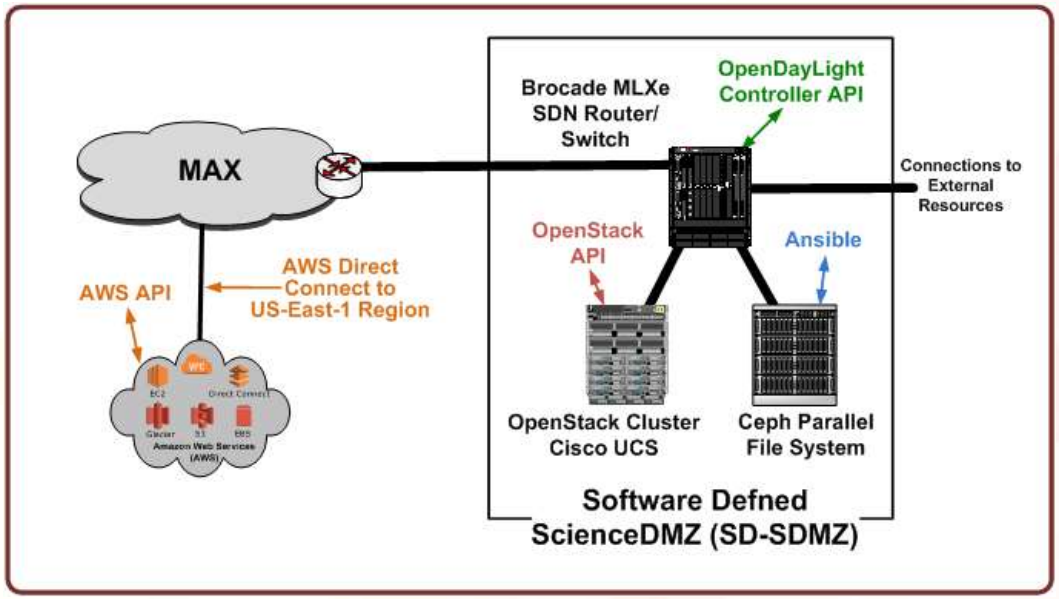
THANKS

MAX SD-Science DMZ Resources

Brocade MLXe:
- OpenFlow 1.3 Capable
- 8x100G Ports
- 4x40G Ports
- 24x10G Ports
- 48x1G Ports

Cisco Unified Computing System (UCS):
- 12 Compute Blades, dual socket, multi-core
- 2x2 redundant Fabric Interconnects with FEX technology
- 2x3548 Nexus OpenFlow capable switch
- running OpenStack Liberty

Ceph (luminous)/Ethernet High Performance File System:
- 6 Object Storage Devices at 36 TB each (12x3TB drives)
- Approximately 200 Terabytes high performance cache storage



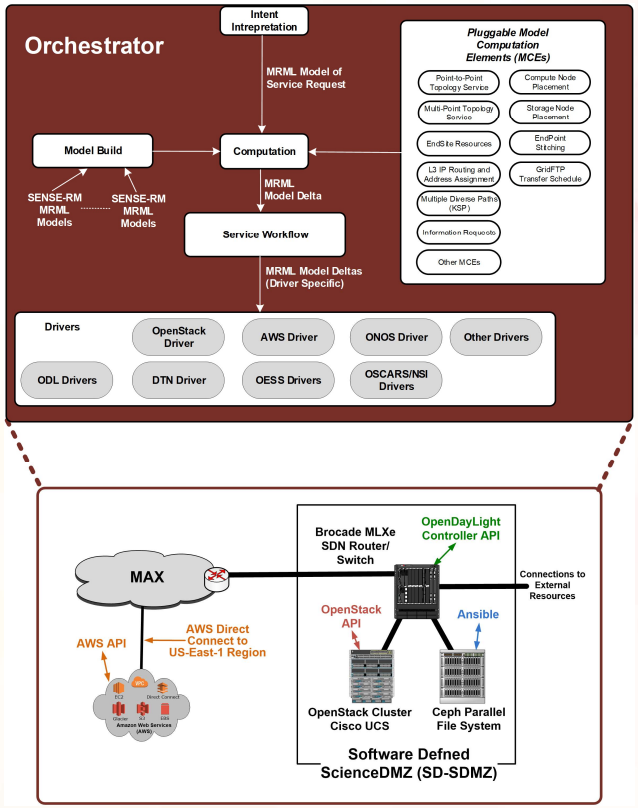
- Each OSD chassis
  - 2U Chassis, 2 Intel Xeon E5-2609 Quad-Core, 2.4Ghz CPUs
  - 8GB Memory
  - LSI MegaRaid 9280-16i4 SAS, 6GB/s PCI-e RAID Card
  - Dual Port 10Gbe NIC card
  - 12 3 Tbyte SATA 6GB/s Hard Drives

Standards/Open Source based User Identity, Authorization, Federation

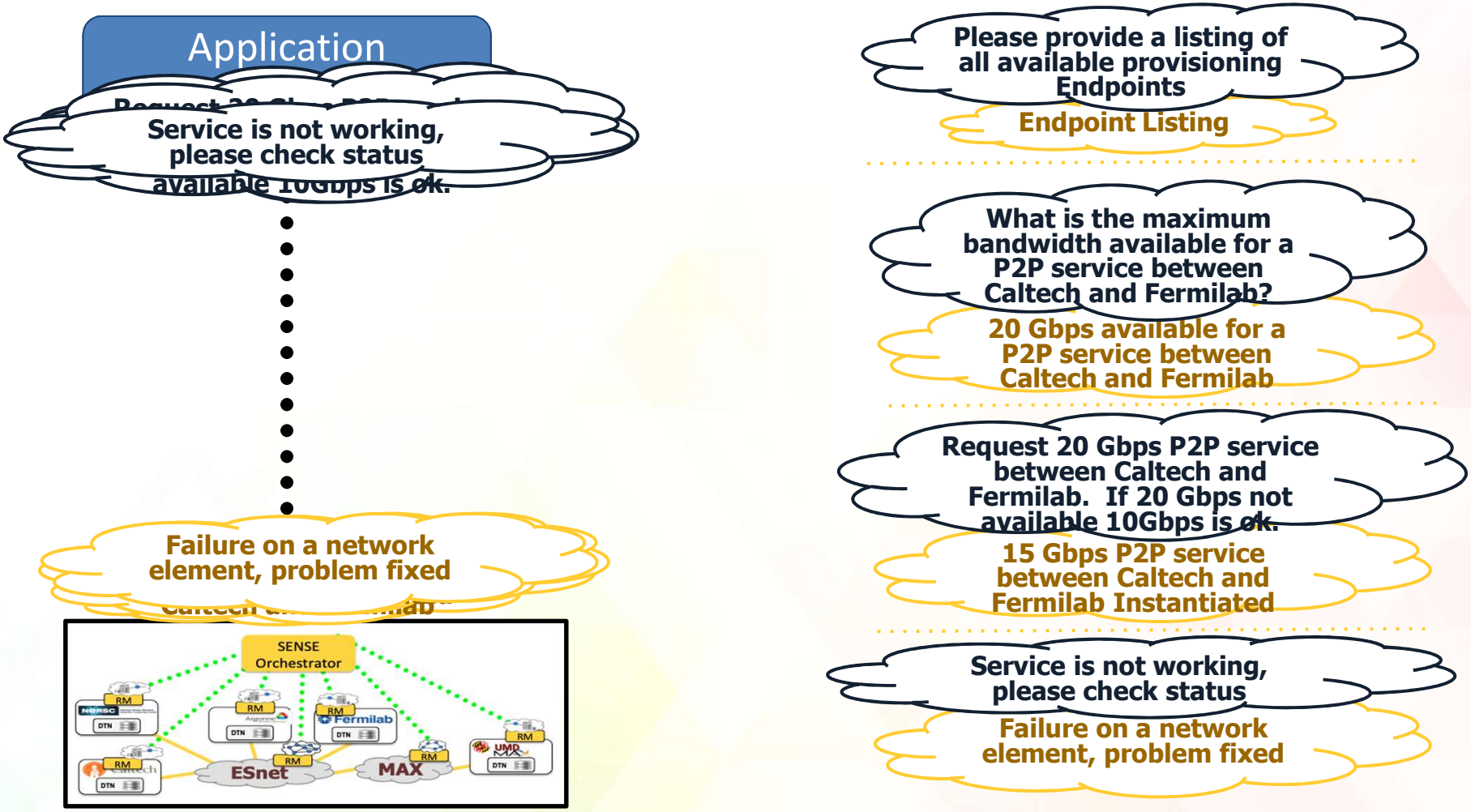
Keycloak
- Role based access for StackV Web Pages/Services
- Single Sign On (OAuth, OpenID)
- Federation (Shiboleth)
- opensource
- www.keycloak.org

FreeIPA
- Local Identity Provider
- HBAC (Host Based Access Control)
- Centralized kerberos management for VMs
- LDAP for MetaData
- Integration with external automation (ansible, others)
- opensource
- www.freeipa.org

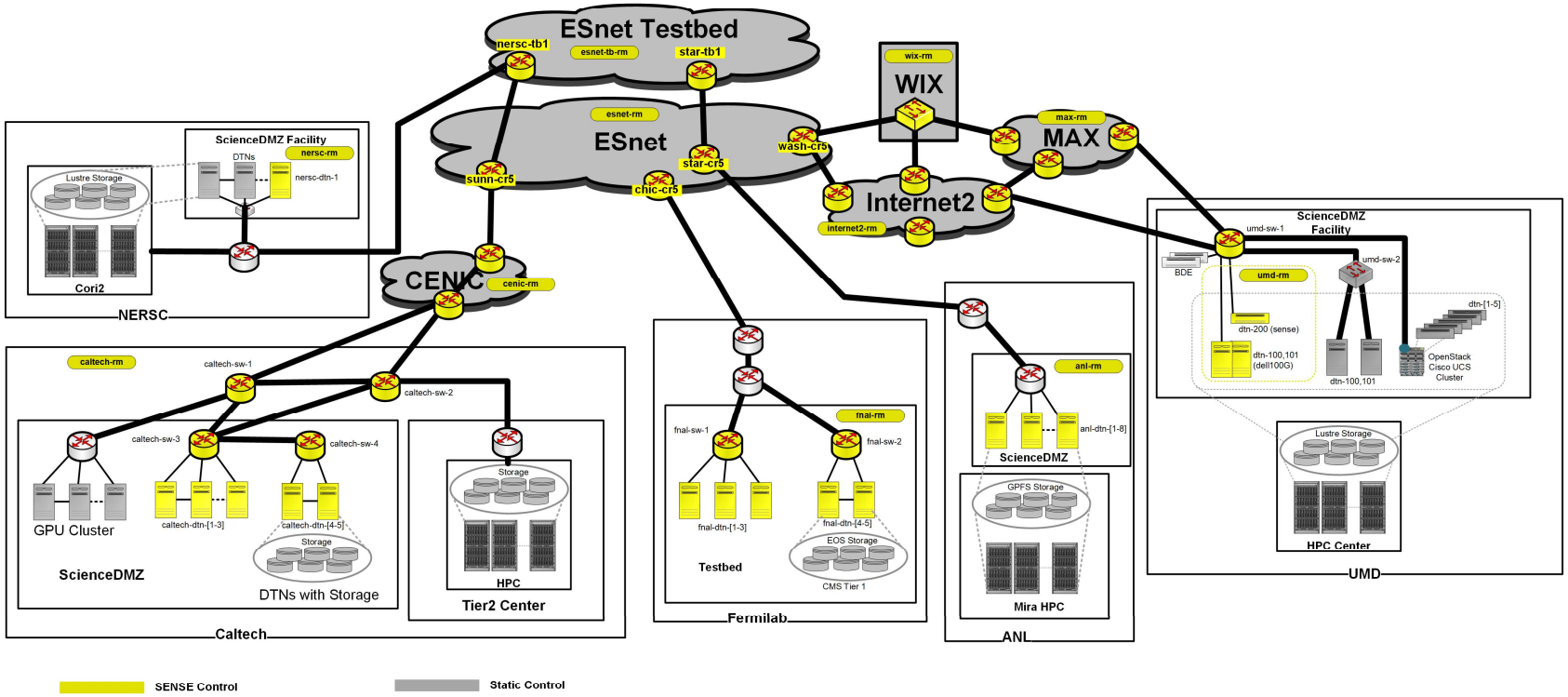
KeyCloak FreeIPA



Application to Orchestrator Interactions



SENSE Testbed Deployment



SENSE Scalability – 68 Domains Test