MAX Edge and Cloud Services

MAX Participants Meeting
April 25, 2018
College Park, Maryland

Tom Lehman
Director of Research
UMD/MAX

Xi Yang
Senior Research Scientist
UMD/MAX

UNIVERSITY OF MARYLAND

MAX
MID-ATLANTIC CROSSROADS

Research Projects

- **UMD/MAX Research Team**
  - Tom Lehman
  - Xi Yang
  - Alberto Jimenez
  - Multiple Students
- **Results from several research projects including:**
  - GENI Enabled Software Defined Exchange (SDX)
  - High Performance Computing with Data and Networking Acceleration (HPCDNA)
  - Regional Embedded Cloud for As-a-Service Transformation (RECAST)
  - Resource Aware Intelligent Network Services(RAINS)

Today's Topics

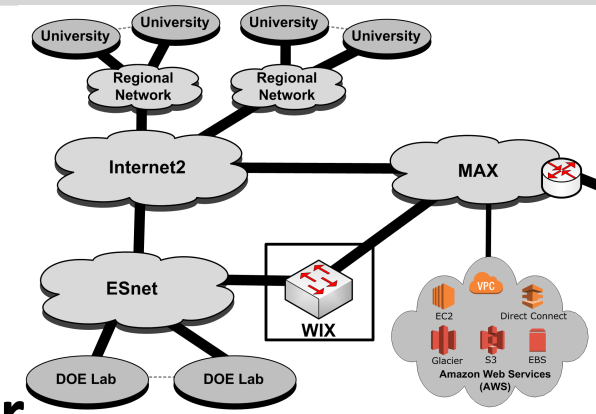
- **Next Generation Cyber Infrastructure "Edge Facilities" vision:**
  - **Software Defined Science DMZ (SD-SDMZ)**
  - **Software Defined Exchange (SDX)**
- **New MAX Pilot Service Announcements**
  - **Advanced Hybrid Cloud Service**
  - **EdgeCloud Connection Service**
    - **MAX and Internet2 Collaboration**

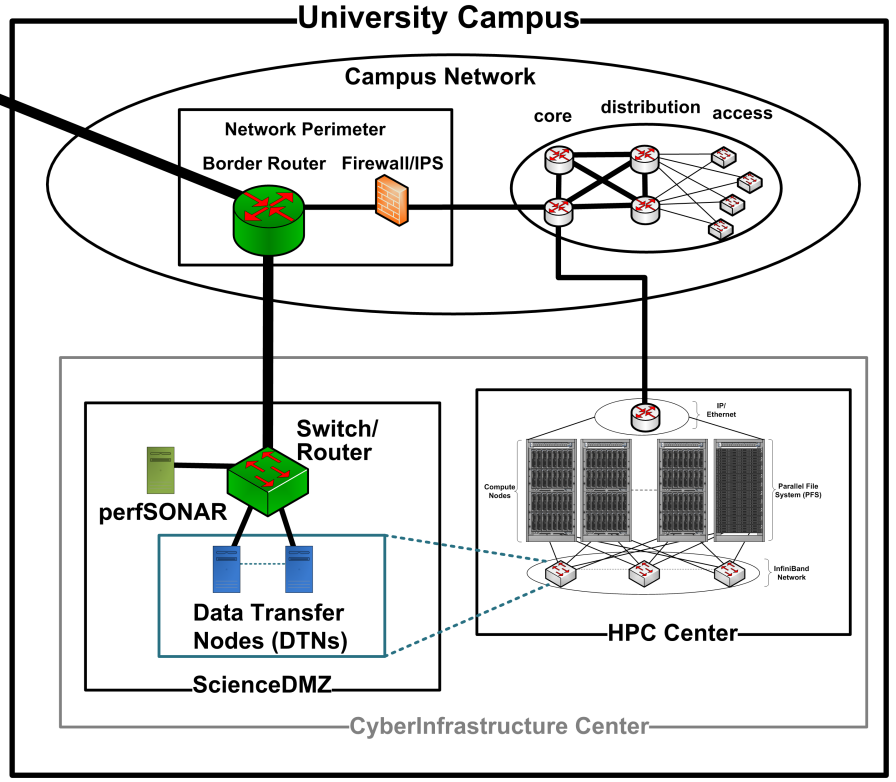
Next Generation Cyber Infrastructure

- **Edge Facilities will be an important component**
  - Located at the edge of an enterprise and/or network
  - May include network, compute, and/or storage resources
  - Software Defined Infrastructure (SDI) based
  - Leverages rich connectivity and programmability to provide new and flexible services
- **We have a specific vision (and deployments) for the following instantiations of "Edge Facilities":**
  - Software Defined Science DMZ (SD-SDMZ)
  - Software Defined Exchange (SDX)

Traditional Science DMZ

- **Bare metal Data Transfer Nodes, PerfSonar**

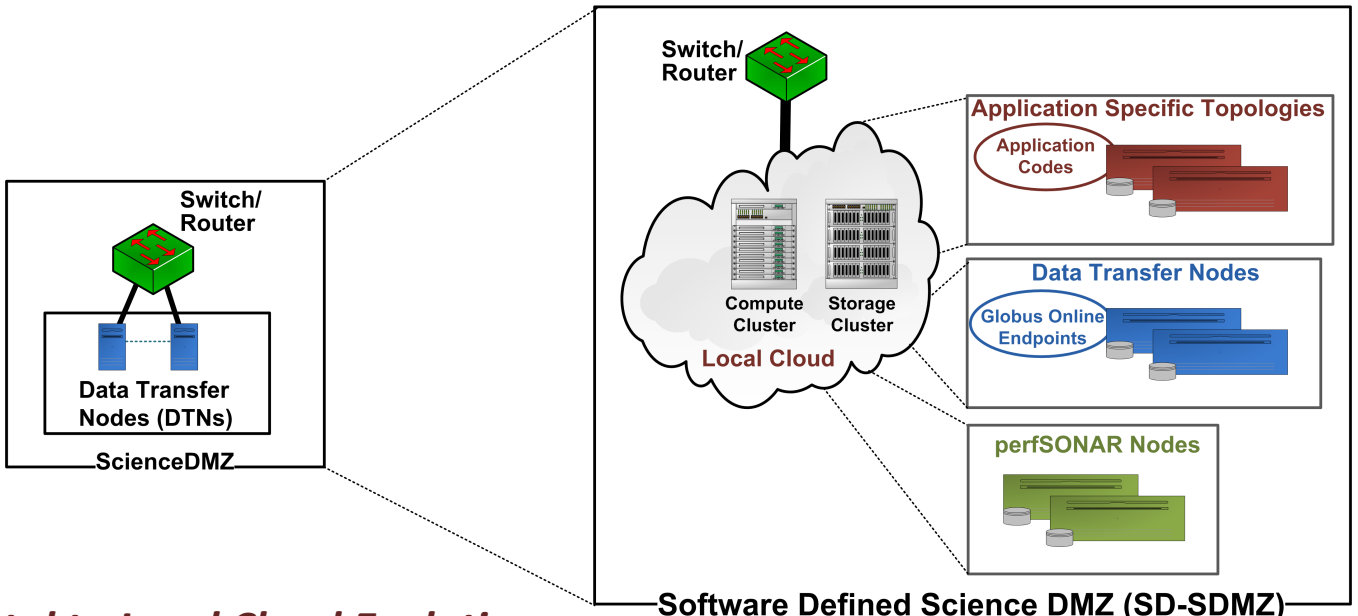




- **Manual network (switch/ router) control**
- **Primary single service focus of Data Transfer (Globus)**

Step 1: Virtualize and Cloudify

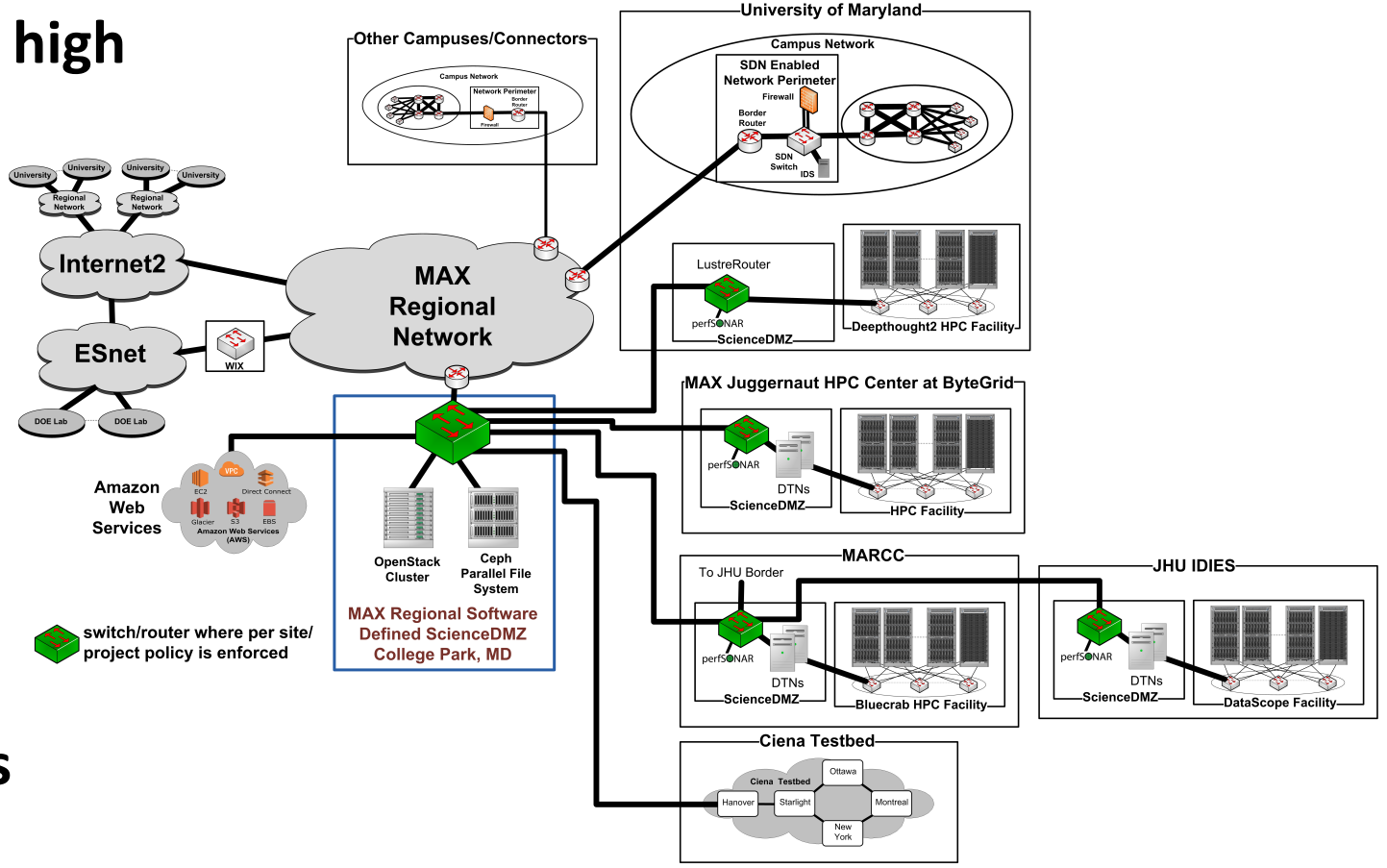
- **Transition to Local Compute Cluster, Storage, SDN Network Control. On Demand, scalable, traditional services and advanced hybrid cloud services**



Bare Metal to Local Cloud Evolution

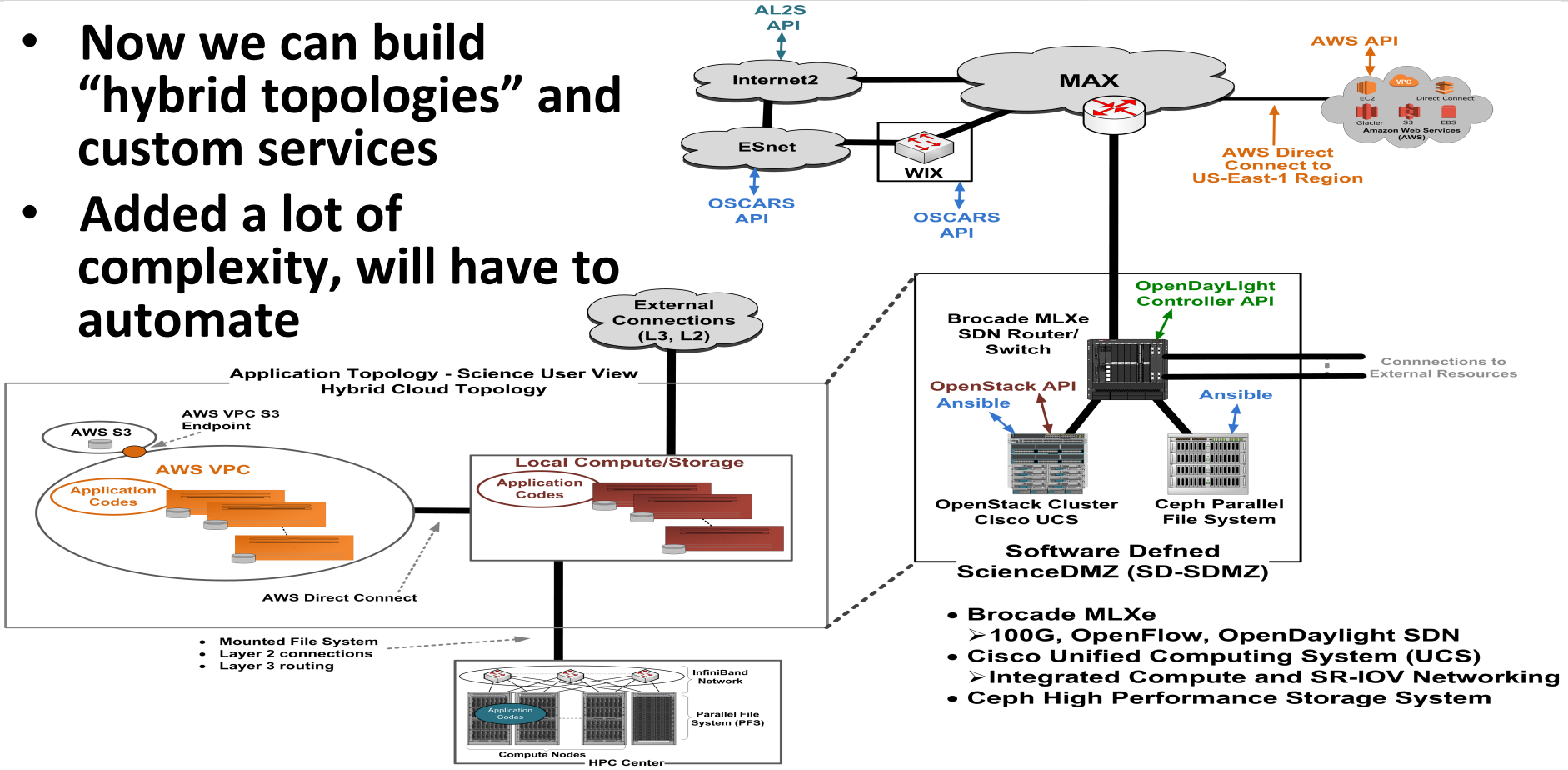
Step 2: Regionalize

- Leverage the high performance connections available via regional networks
- Transition to “SDMZ as a Service” for multiple organizations



MAX SD-SDMZ ("Protoduction Facility")

- Now we can build "hybrid topologies" and custom services
- Added a lot of complexity, will have to automate

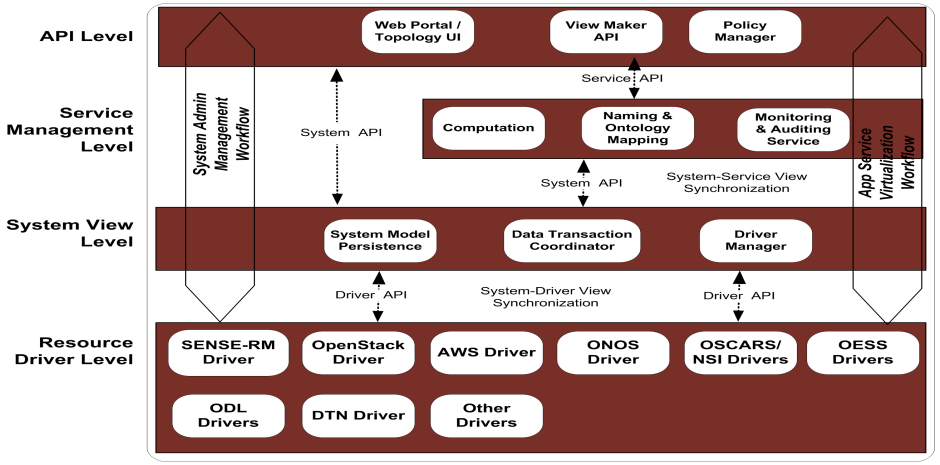


- Brocade MLXe
  - ➢100G, OpenFlow, OpenDaylight SDN
- Cisco Unified Computing System (UCS)
  - ➢Integrated Compute and SR-IOV Networking
- Ceph High Performance Storage System

Our Approach and Solution

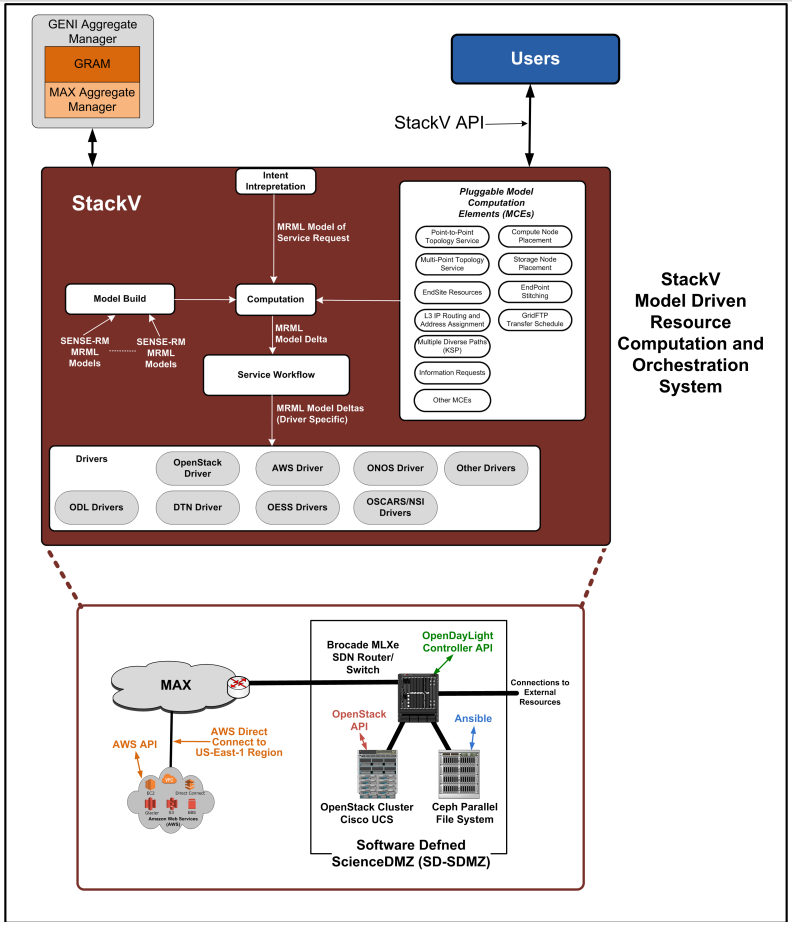
- **Multi-Resource Orchestration:** integrating and orchestrating the network and network services with the things that attach to the network – compute, storage, clouds, and instruments.
- **Model Driven:** using models to describe resources in order to allow integrated reasoning, abstraction, and user centered services
- **Intelligent Computation Services:** Model driven, multi-resource computation services to enable orchestration services in response to high level user requests.
- We want to "Orchestrate the Automaters"

StackV - Orchestration Suite



Architecture

Conceptual View

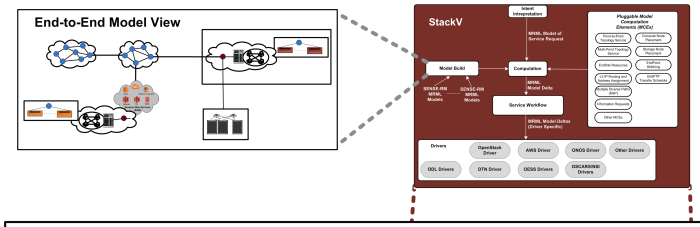


- UMD/MAX developed StackV is an open source (geni public license) model driven orchestration system:
  - github.com/MAX-UMD/stackV.community
- Would like to build an open source community around, available for others to use as desired

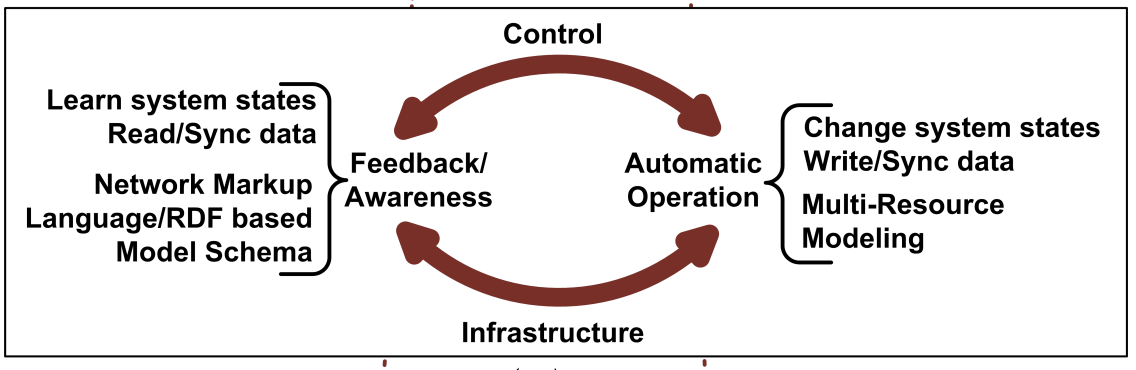
Model Based Control and Orchestration

StackV Design Principles

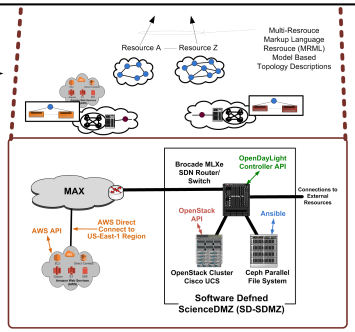
- Focus is turnkey services which can be customized for different users
- Model based allows for rapid integration of new resource types
- Modular Computation Elements facilitate custom service workflow construction
- A DevOps model for service construction
- Generic control of infrastructure for custom construction of user facing services



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

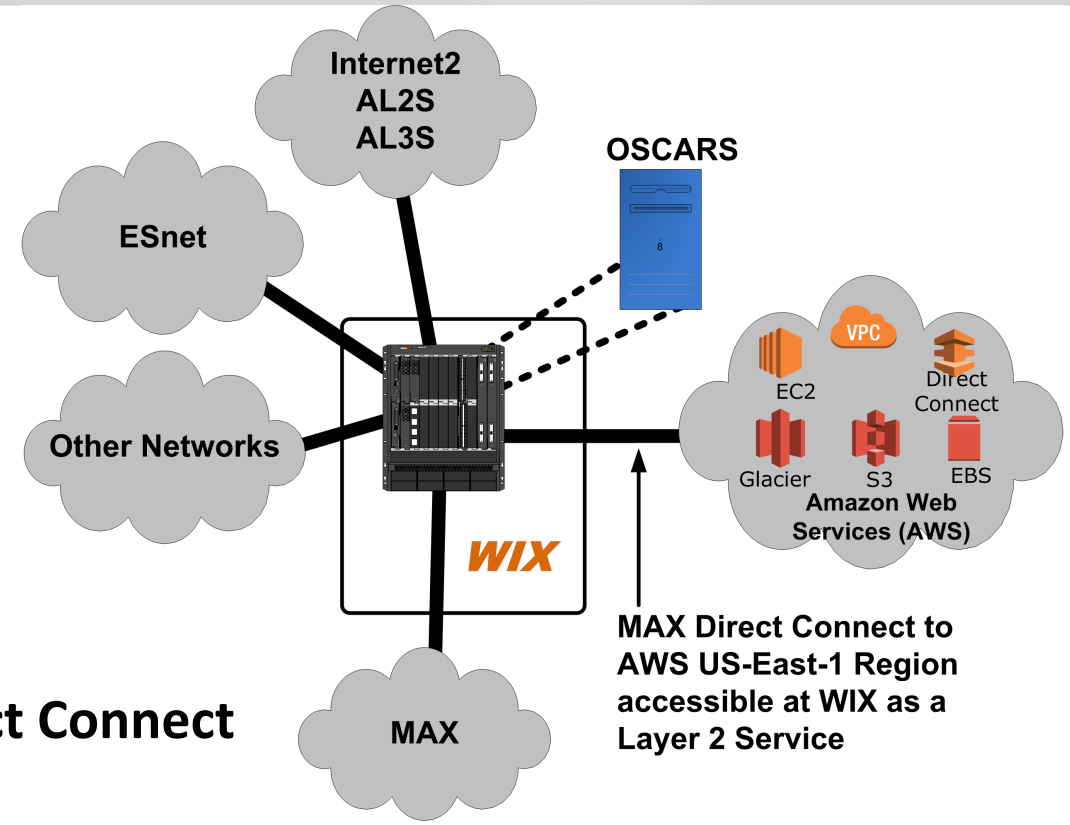


Automated Infrastructure Models created using available System APIs



Software Defined Exchange (SDX)

- Same technology base can be used for SDX
- WIX is a production Exchange Point in McLean, Virginia (jointly operated by Internet2 and MAX)
- Includes OSCARS service enabling Dynamic Cross Connects
- MAX has made its AWS Direct Connect Service available at the WIX

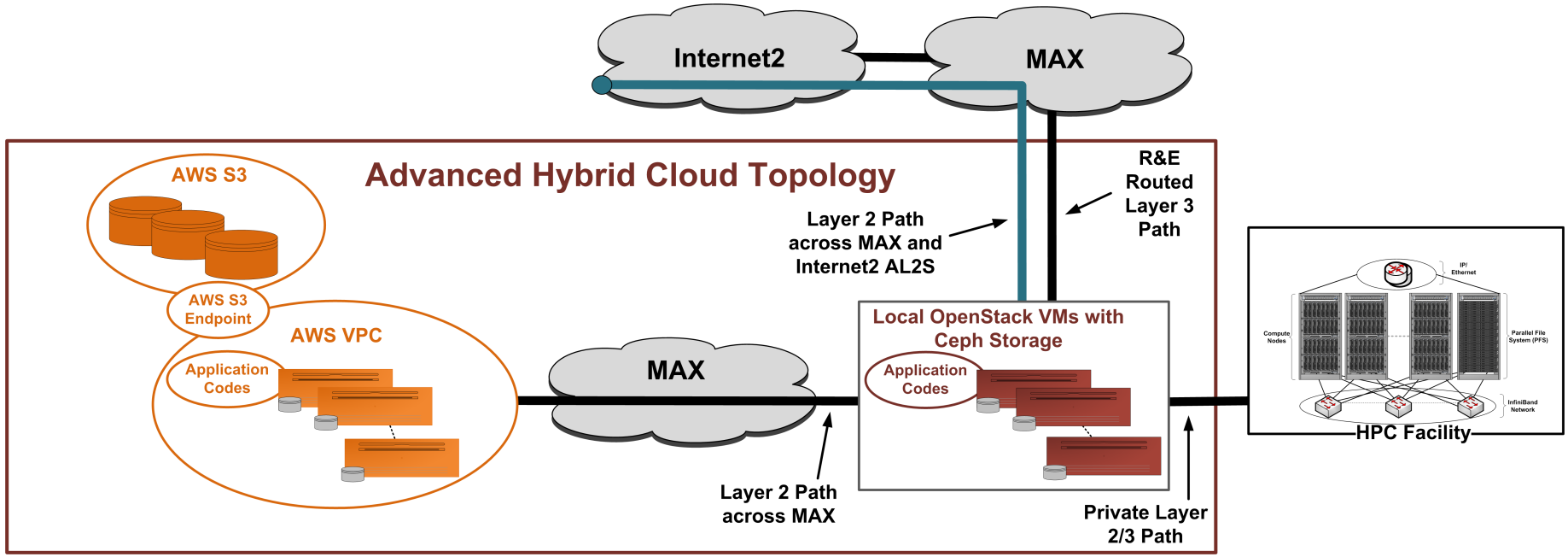


Pilot Services

- **Advanced Hybrid Cloud Service**
- **EdgeCloud Connection Service**
- **Service Objectives**
  - **Work with users to determine what is useful**
  - **Modify and refine the services, and identify other services that would be valuable**
  - **Develop a plan to transition from protoduction/pilot service to production service**
- **Pilot Service Timeline: May – September 2018**

Advanced Hybrid Cloud Service

Example Topology



SD-SDMZ Services

Advanced Hybrid Cloud (AHC) Service

On Demand, Application Specific, Hybrid Topologies which include one of more of the following:

- ✓ **Local OpenStack Virtual Machines (with SRIOV interfaces to network and storage)**
- ✓ **Dedicated Local Ceph Storage Resources and Connections**
- ✓ **Integrated AWS Resources (Virtual Private Cloud (VPC) or Public)**
  - **User controlled AWS resources, or**
  - **SD-SDMZ facility provided AWS resources (EC2, Storage, S3 endpoints)**
- ✓ **Network Connections**
  - **AWS Direct Connect integration for access to AWS Public or VPC resources**
  - **Layer2/Layer2 External Connections across Internet2, ESnet, others**
  - **Service connections/integration with other R&E cloud, HPC, data repositories, etc.**
  - **Customized topology templates for individual user requirements**
- ✓ **Future:**
  - **Service connections/integration with other public cloud infrastructures**
  - **Schedulable Services**

Scheduling and QoS

- We are “orchestrating” automated resources
  - which sometimes do not have scheduling or QoS features
- Therefore we cannot provide scheduling and QoS for all resources
- Approach
  - put those resources under Schedule and QoS control that we can AND when it adds value to do so
- The AWS Direct Connect Link is one resource where we think this adds value

Scheduling and QoS

- **Two modes for use of the AWS Direct Connect Link:**
  - **Shared (with preemption)**
  - **Dedicated (Bandwidth with Scheduling)**
- **Shared Mode**
  - **Multiple users on best effort basis, may be preempted**
  - **Preemption results in AWS Virtual Circuit VLAN being removed from SD-SDMZ or SDX network element for a period of time. This will interrupt the AWS BGP session.**
  - **AWS BGP restoration is automatic upon AWS Virtual Circuit VLAN restoration. No action required by user to restore service.**

Scheduling and QoS

- **Dedicated Mode**
  - **User can request a specific time slot for dedicated use of AWS Direct Connect Link.**
  - **Initially this is for the entire 10 Gbps of bandwidth. Studying options to offer more granular bandwidth options.**
- **Preemption Policy**
  - **In order to share Direct Connect Link resources between Shared and Dedicated users, the following polices are defined. These may be adjusted based on experience and user feedback.**
  - **Shared mode users will receive 2 hours notice (automated email from StackV) prior to preemption**
  - **Total Dedicated Mode usage will be limited to no more than 50% of wall time on a daily basis**

Note: AWS Direct Connect Scheduling and QoS feature is planned for release in July 2018

Scheduling and QoS

- **Use Cases**
  - **This mode of Shared (with preemption) and Dedicated usage is geared toward domain science users or special purpose enterprise operations such as regular backup or recovery operations**
  - **Not focused on regular 24x7 enterprise cloud access and usage model**
  - **The idea is that there may be multiple direct connects to various cloud providers, which are managed in different ways for different use cases.**

Advanced Hybrid Cloud Service

- One time coordination activity with MAX team
  - build customized Service Template
- Subsequently
  - user can independently instantiate and delete service topology
- Service Template will generally be read only
  - but some parameters will be opened up for independent user modification as we get more experience with testing

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 StackV interface)
   - AL2S (remote endpoint)
   - Other external resources as available

4) AHC Service Level Parameters
   - Define number of simultaneous service instances allowed

Advanced Hybrid Cloud Service

- The exact topology of an AHC service can be customized per the user requirements
- The orchestration system is actually constructed to allow users to independently discover available resources and construct their own template. At this point we are planning to implement a more restrictive policy.

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 StackV interface)

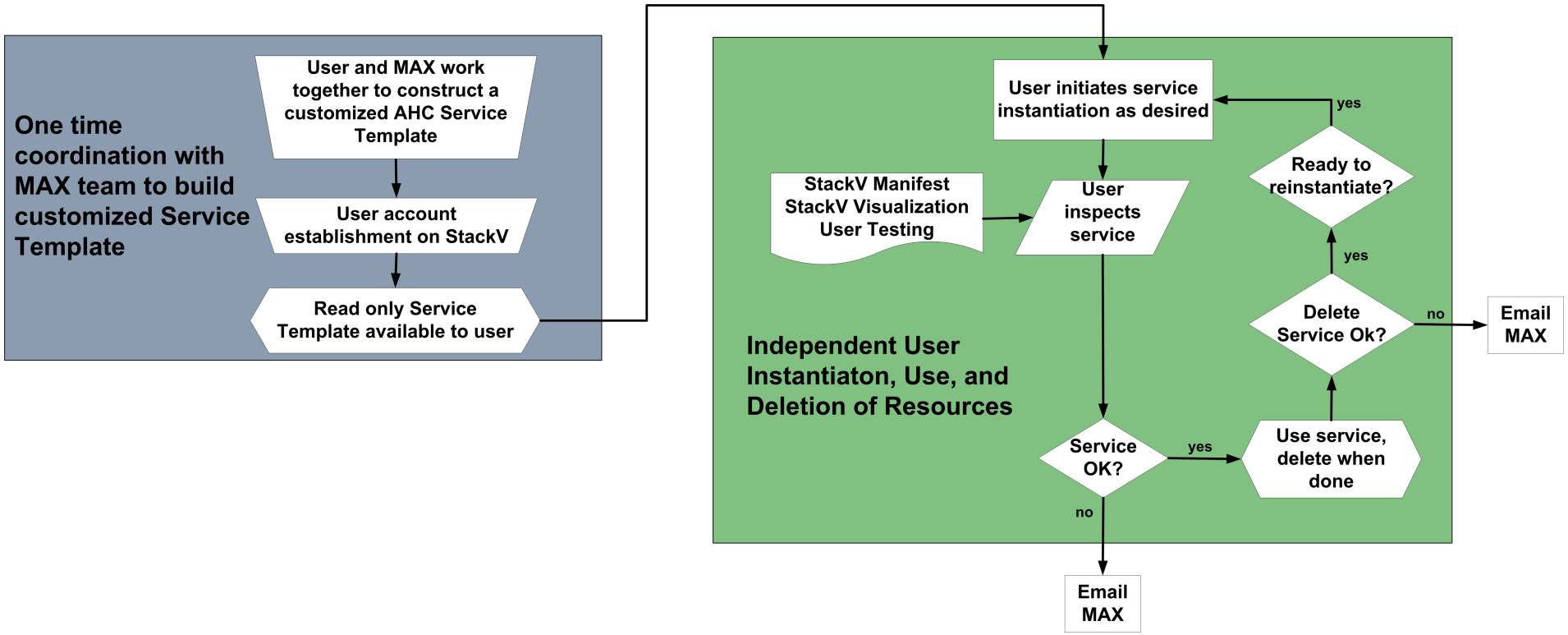
- AL2S (remote endpoint)
- Other external resources as available

4) AHC Service Level Parameters

- Define number of simultaneous service instances allowed

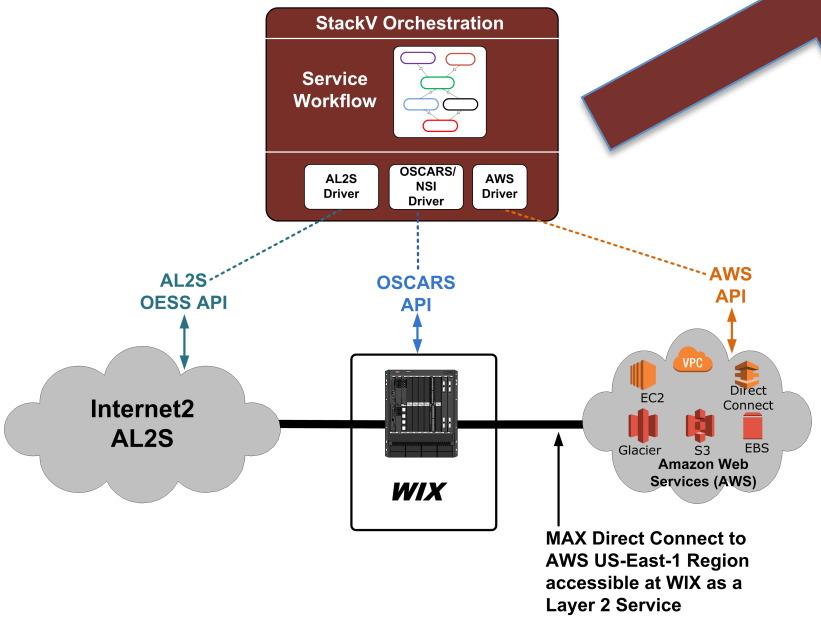
Advanced Hybrid Cloud Service

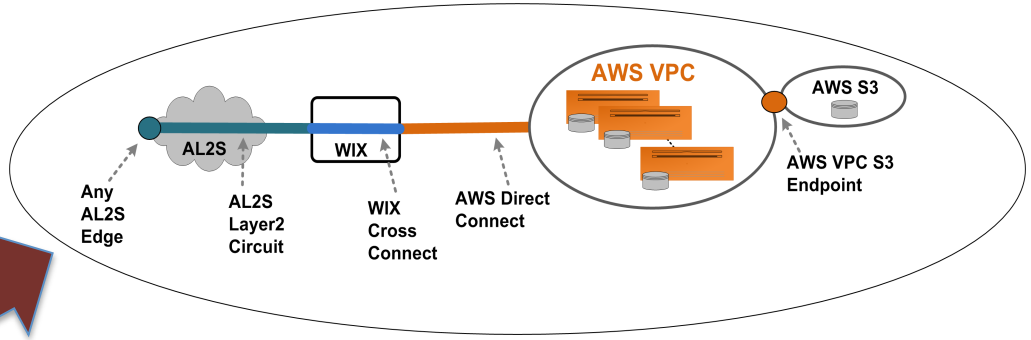
User Workflow



EdgeCloud Connection Service

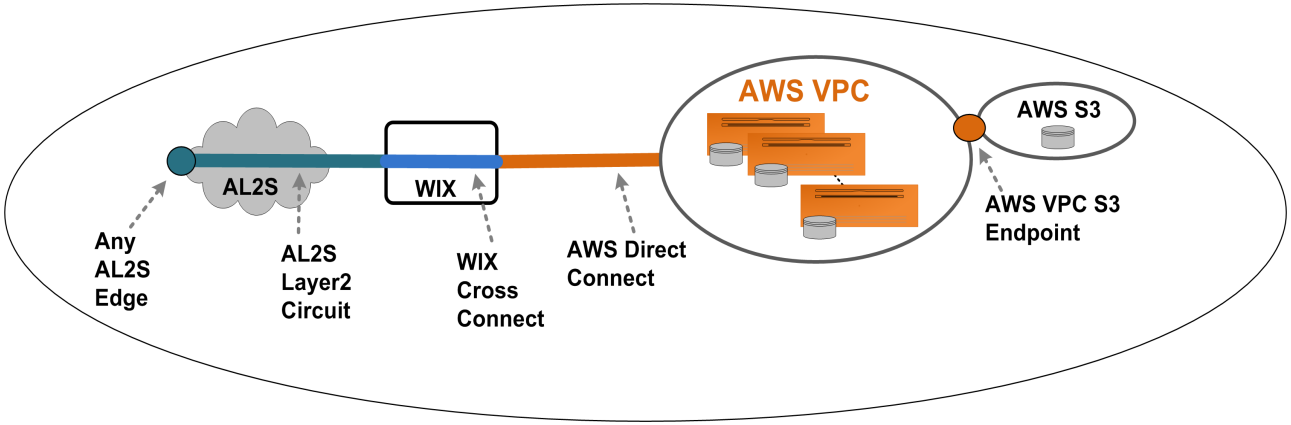
Separate StackV instance for this service





Orchestrates AL2S, WIX, and AWS Direct Connect provisioning

EdgeCloud Connection Service



- Same workflow as for other service
- Different template
- Not available: compute, storage, bgp instance for AWS peering
- Available: AWS Direct Connect via shared or scheduled/dedicated mode

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

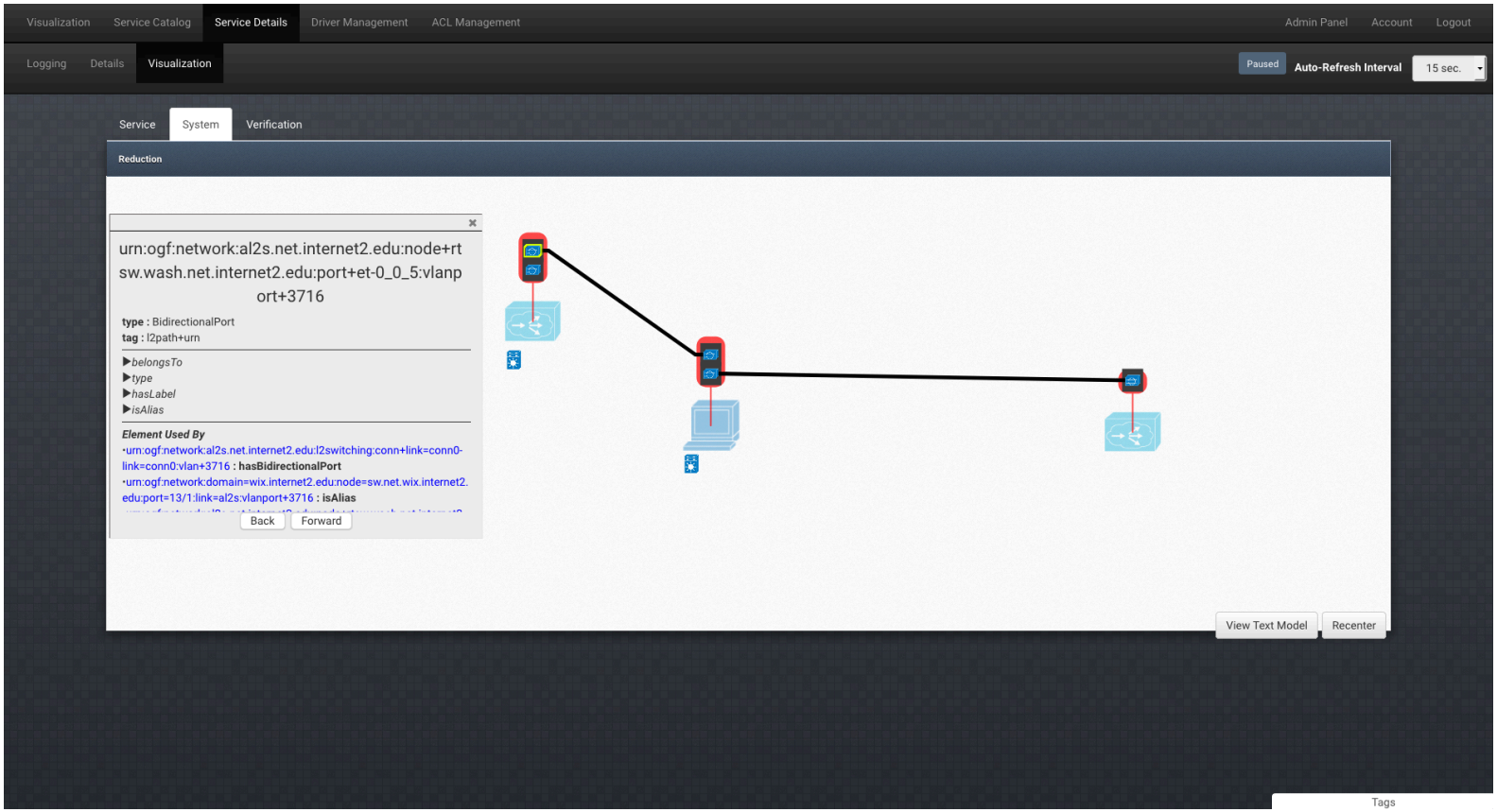
2) External Connections

- AL2S (remote endpoint)

3) AHC Service Level Parameters

- Define number of simultaneous service instances allowed

StackV - EdgeCloud Connection Service Topology



Pilot Service Participation

- **If interested in participating in pilot, or obtaining more information send email to:**
  - **research at maxgigapop dot net**
- **SD-SDMZ Advanced Hybrid Cloud (AHC) Service Demonstration Video:**
  - **https://tinyurl.com/max-sdmz-pilot**
- **EdgeCloud Connection Service Demonstration Video:**
  - **https://tinyurl.com/max-ecc-pilot**

Thanks

Extras

MAX SD-Science DMZ Resources

Brocade MLXe:

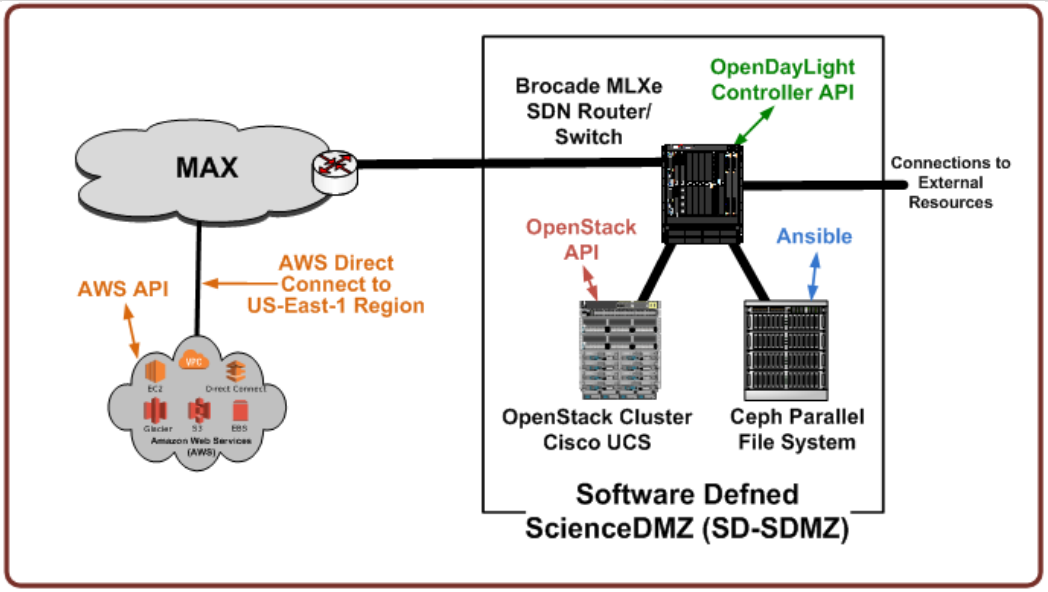
- OpenFlow 1.3 Capable
- 8x100G Ports
- 4x40G Ports
- 24x10G Ports
- 48x1G Ports

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

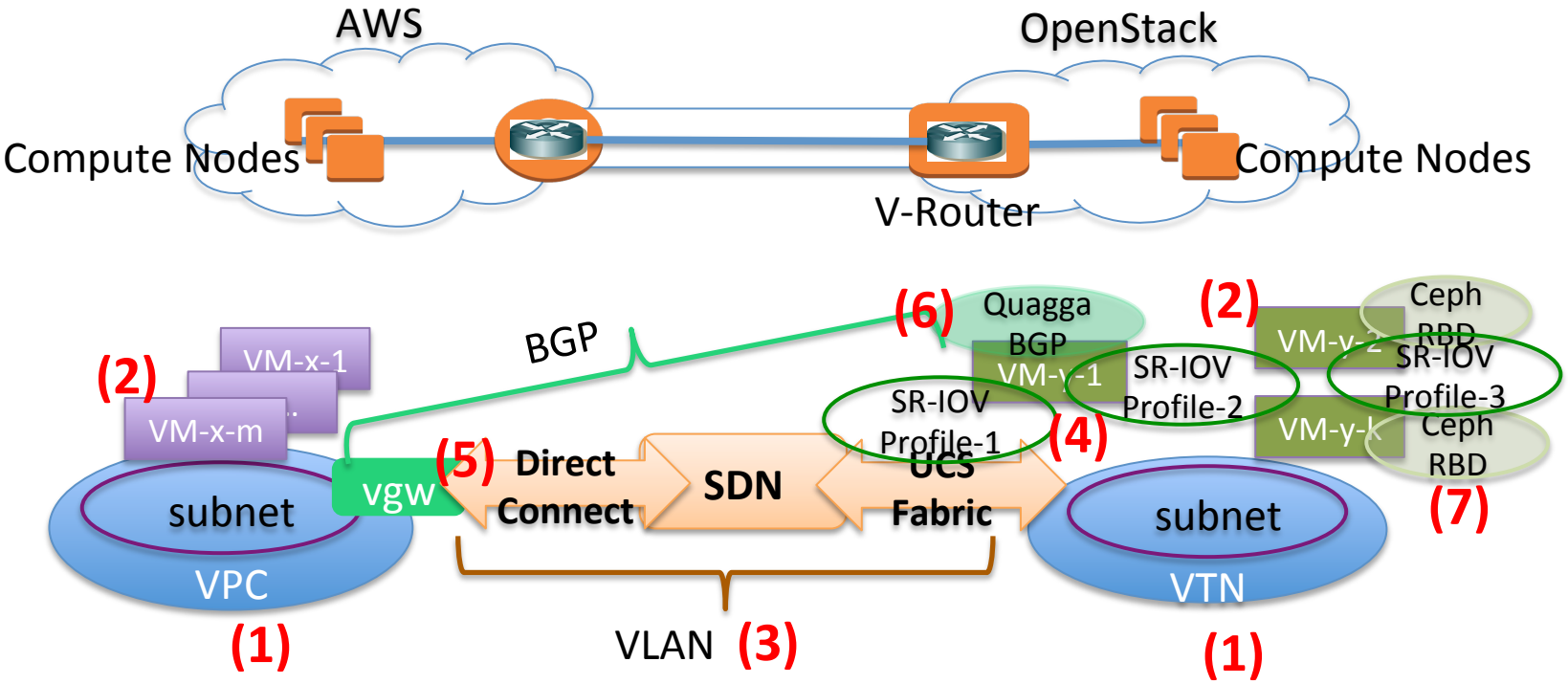


- Each OSD chassis
  - 2U Chassis, 2 Intel Xeon E5-2609 Quad-Core, 2.4Ghz CPUs
  - 8GB Memory
  - LSI MegaRaid 9280-16i4 SAS, 6GB/s PCI-e RAID Card
  - Dual Port 10Gbe NIC card
  - 12 3 Tbyte SATA 6GB/s Hard Drives

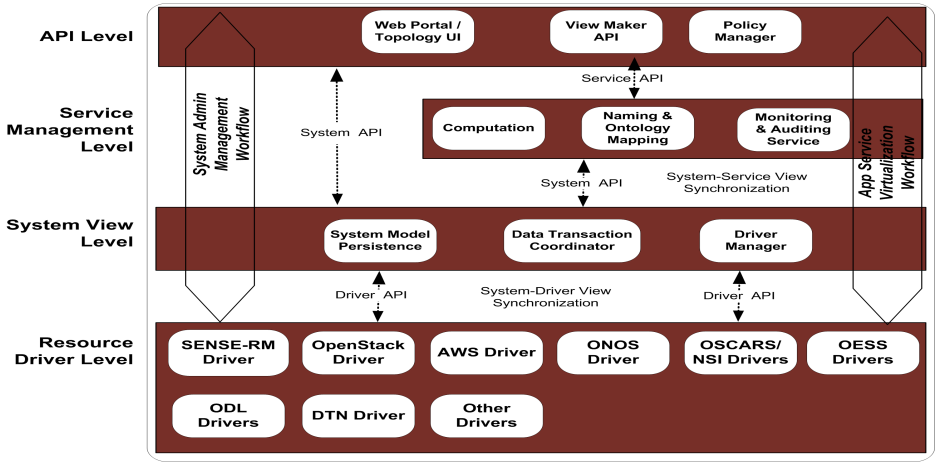
Model Driven Orchestration

- Modeling schemas based on OGF Network Markup Language (NML).
- Developed extensions to allow resources which are connected to the network to also be modeled: Multi-Resource Markup Language (MRML)
  - https://github.com/MAX-UMD/nml-mrml
- Utilizes Resource Description Framework (RDF) and OWL 2 Web Ontology Language W3C Specifications

SD-SDMZ Demonstration
Advanced Hybrid Cloud Orchestration



StackV - Orchestration Suite



- API Level - Web Portal, Northbound Interface for Client Interface
- Service Management Level - Computation Engine, Workflow Management
- System View Level - Unified topology view construction, Driver management
- Resource Driver Level - Southbound Interface to interact with resources

- UMD/MAX developed StackV is an open source (geni public license) model driven orchestration system:
  - github.com/MAX-UMD/stackV.community

- Client Side: Ajax, Jquery, JSON
- API Level: PHP, JSON, Drupal, SOAP, REST
- Service Management Level: Java, C, C++ software, RDF/OWL based schemas
- System Integration Level: Wildfly (Jboss) Java Enterprise, Java Beans
- Resource Driver Level: Java, Python

Standards/Open Source based User Identity, Authorization, Federation

Keycloak

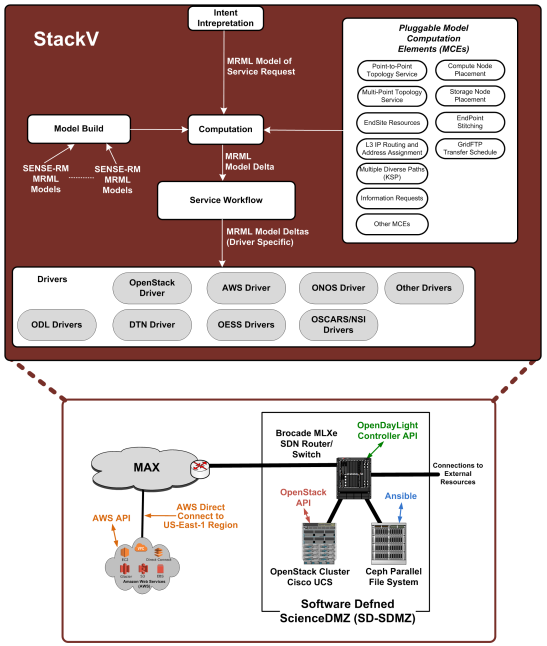
- Role based access for StackV Web Pages/Services
- Single Sign On (OAuth, OpenID)
- Federation (Shiboleth)
- opensource
- www.keycloak.org

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



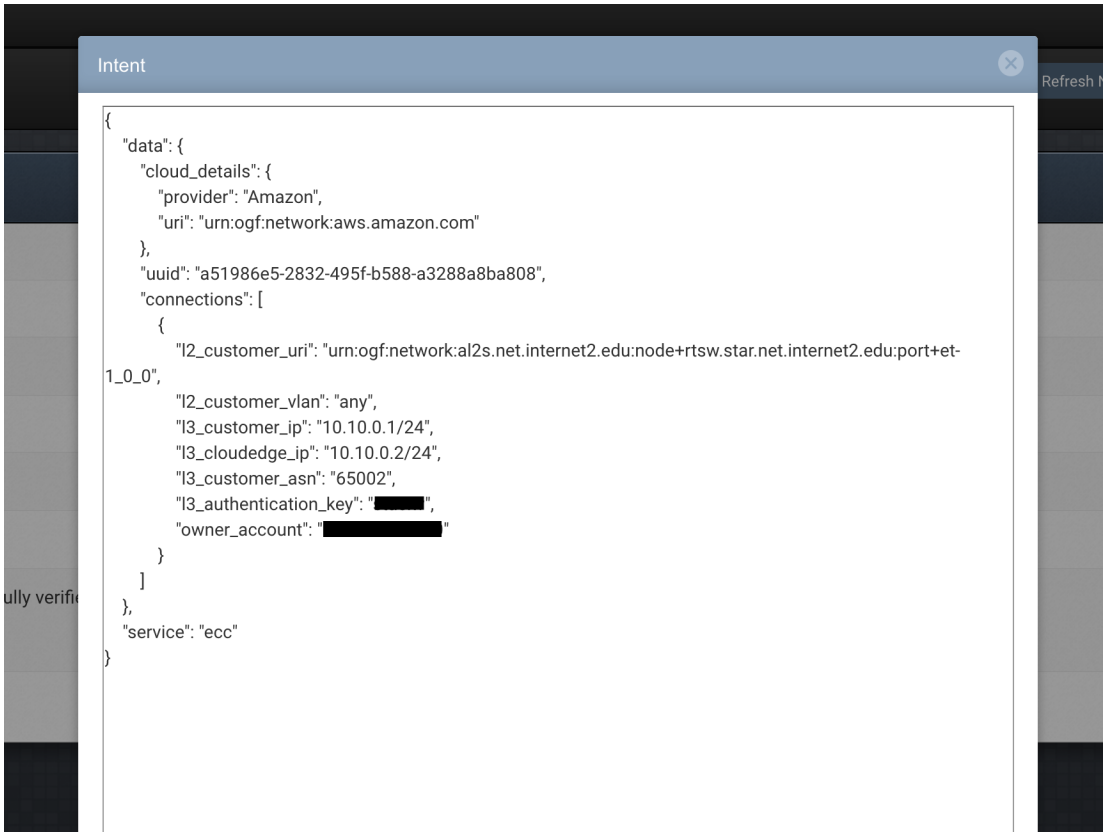
StackV Service Provisioning

- **Screenshots**
  - Initial Request Template
  - User level visualization of verified service
  - Manifest Text
  - Screen for instantiation, delete, showing licenses/allocations

StackV - Intent Request

EdgeCloud Connection Service

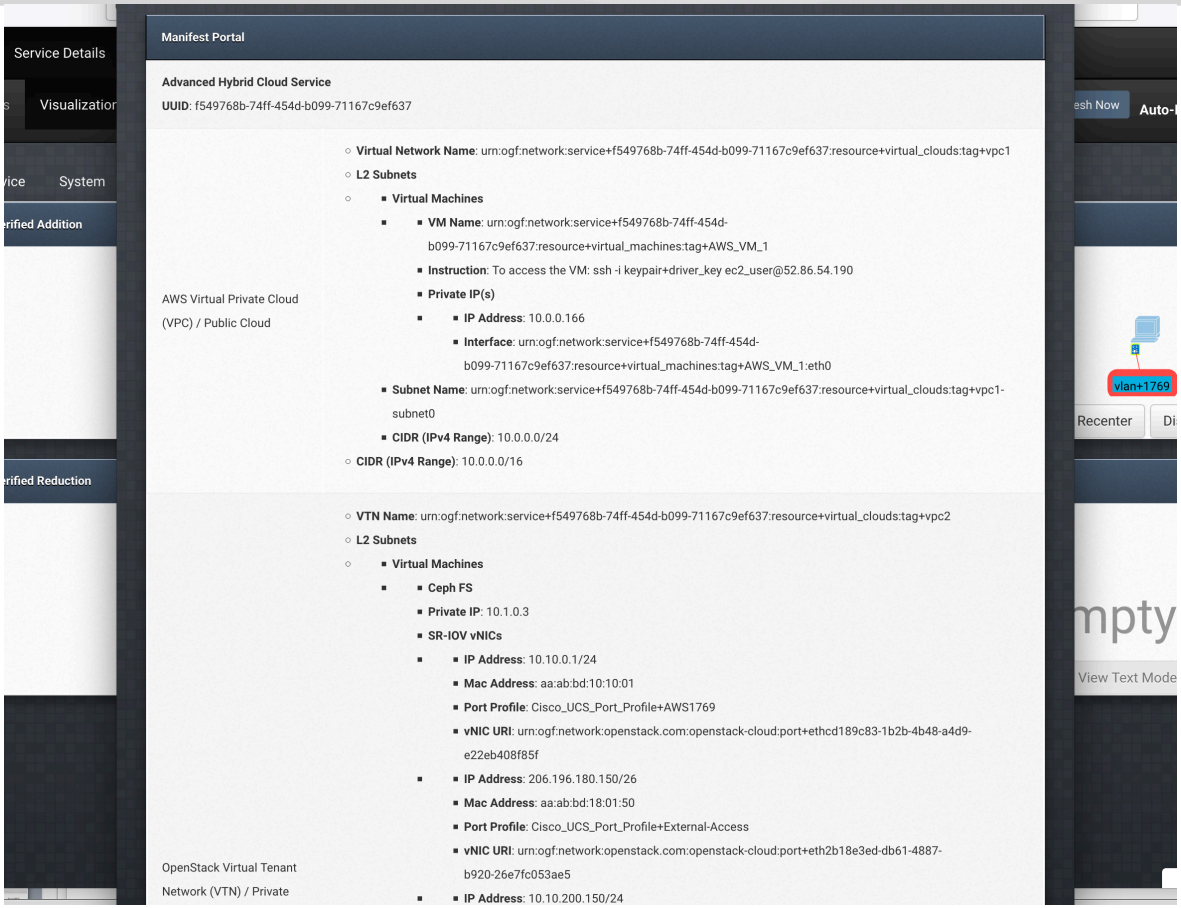
Definition of User Request



StackV – AHC Service Manifest

AHC Service Manifest

Provides Service Details



Example SD-SDMZ Use Cases

- **Global Land Cover Facility (GLCF)**
  - Hybrid cloud topology to facilitate data download from R&E and AWS S3 locations to local HPC filesystem
- **Pan-STARRS Astronomy**
  - Local compute/storage resources to facilitate download and inline processing of telescope data
- **Large Scale Geosimulation**
  - Hybrid cloud topology to facilitate Hadoop cluster set up with local nodes and scalable bursting in to AWS

Future Activities

- **Additional Cloud Provider Support**
  - **Microsoft Azure**
  - **Google Cloud Platform**
- **Work with Internet2 Cloud Connect Initiative**
  - **Possibly orchestrate access to remote cloud locations**