RESOURCES & DATA WG AND EXPEDITIONS UPDATES

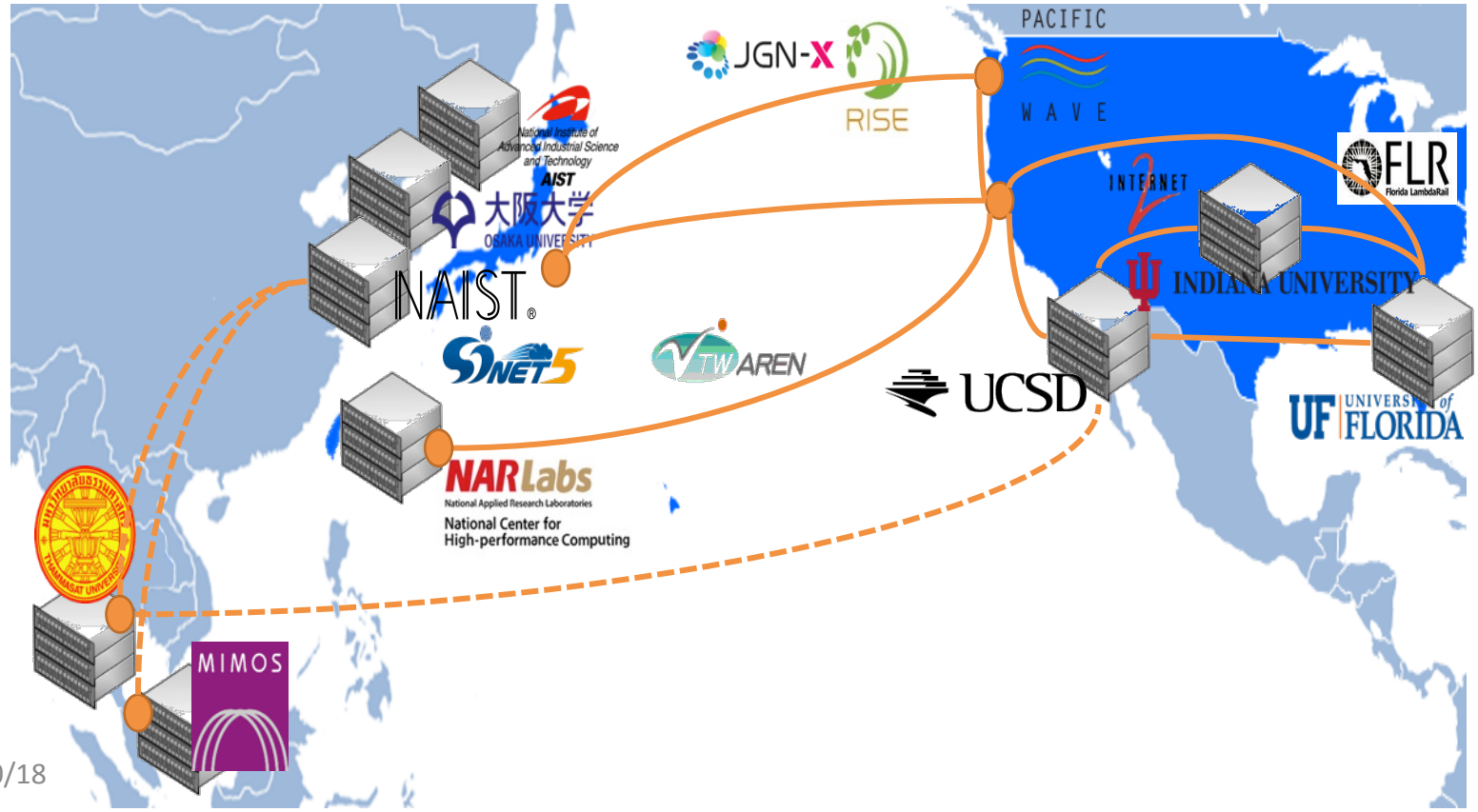
Nadya Williams (UCSD)
Hsiu-Mei Chou (NCHC)

What we planned during PRAGMA 33

- S3 Storage
  - What to use, deploy experimental 2 site
  - What data to store (Airbox, VM, other)
  - Integrate with PRAGMA-ENT
  - Build recipes
  - Evaluate performance
  - Identity management
- Containers
  - Currently run
  - AI images
  - Share “how to”
- Virtual GPU
  - Image analysis
  - Traffic monitoring
- SDN
  - ENT-enabled virtual cluster
  - IPOP

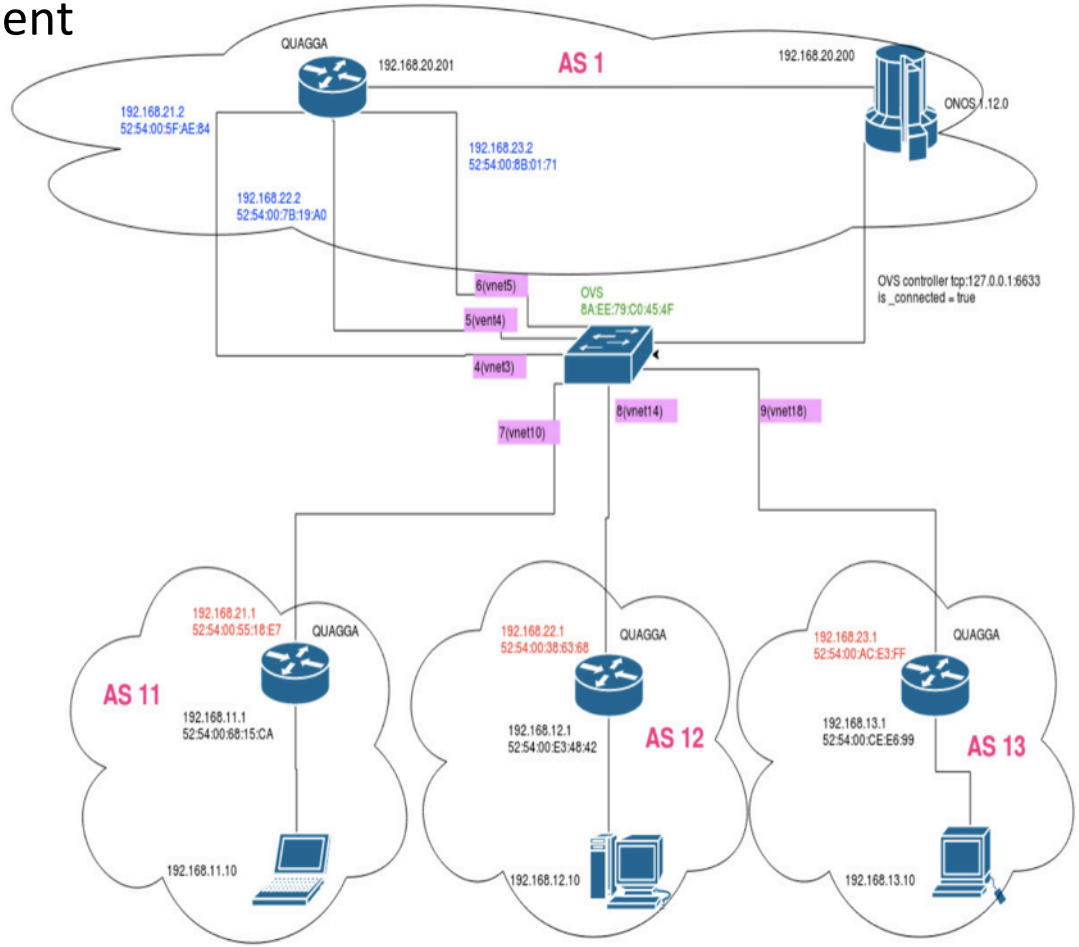
NAIST: Updates of PRAGMA-ENT

- Infrastructure
  - SINET (another Japanese NREN) has been connected to ENT. The connection is dynamically created with NSI (Network Service Interface) in a on-demand manner.



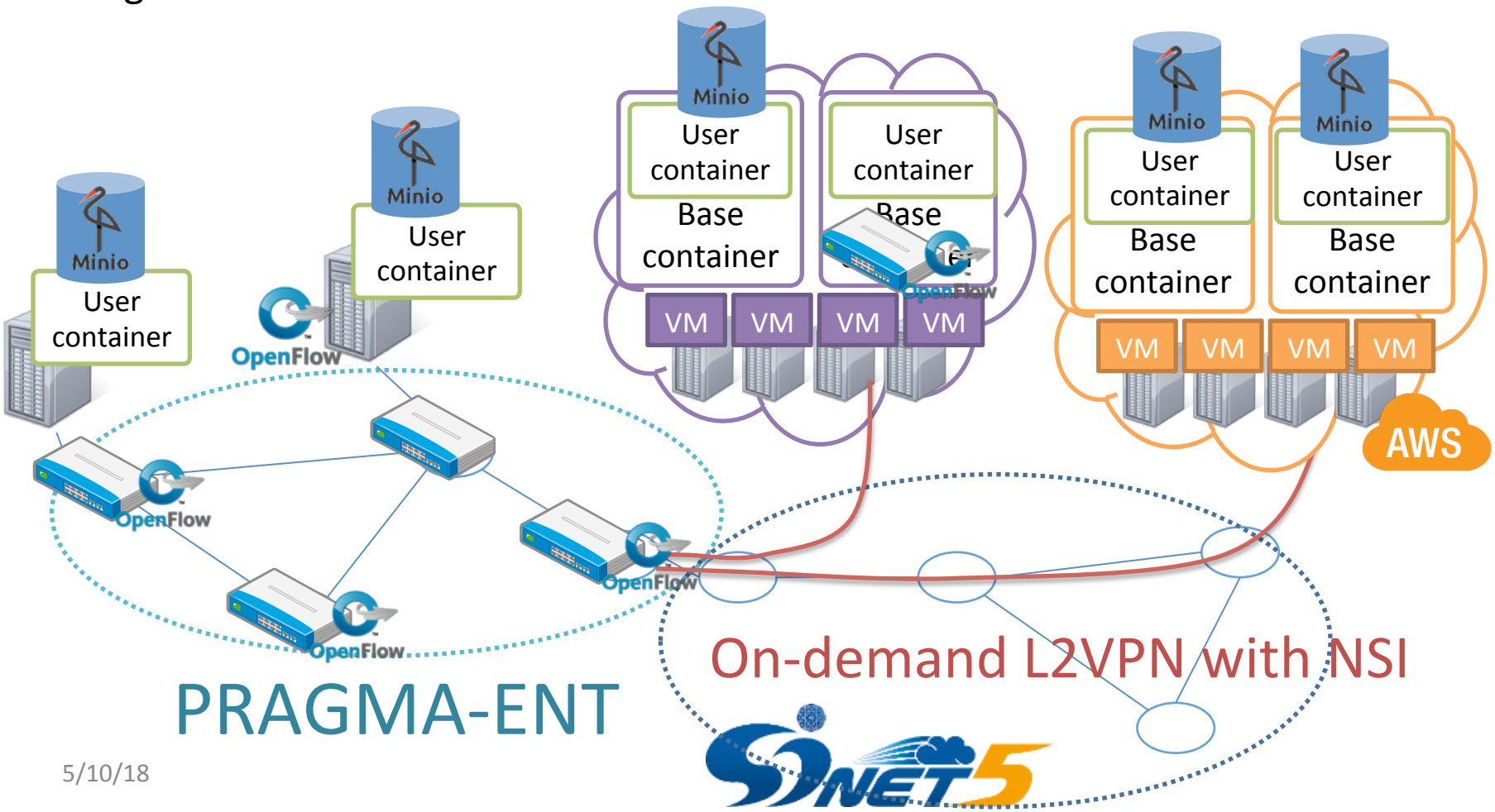
NAIST: Applications update (1): On-demand NFV testbed on ENT and SDN-IP deployment

- pragma_boot can dynamically deploy VMs connected to ENT and create a Network Functions Virtualization (NFV) testbed.
- Sebastian (Osaka Univ.) will present a poster deploying SDN-IP on this environment

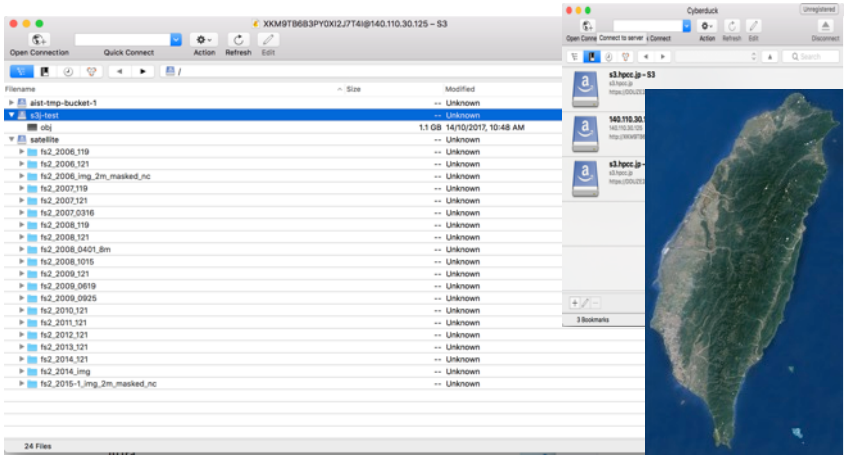


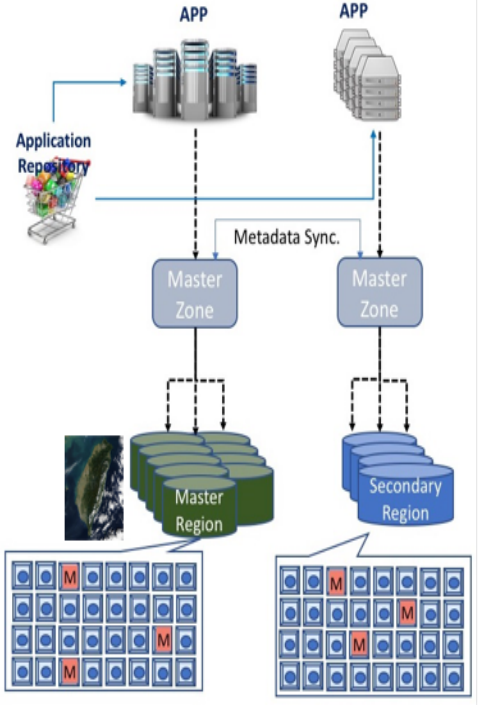
NAIST: Applications update (2): Dynamic storage deployment over the dynamic VLAN service (NSI)

Extends ENT backbone dynamically using Network Service Interface and deploys storage services

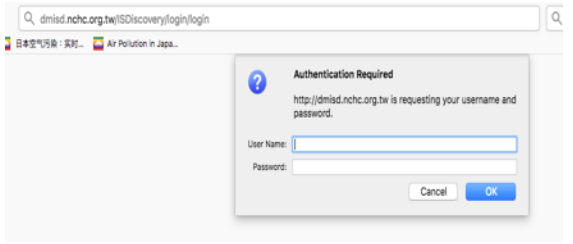


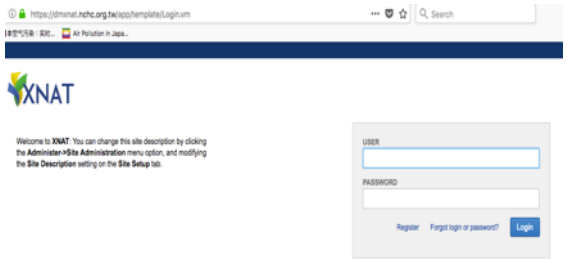
NCHC update





- **S3 system test**
  - Using ceph
  - Provide s3 type of storage connection
  - AIST-NCHC joint collaboration project
- **Airbox data**
  - Stored in DataMart
  - Not ported to S3 yet (sorry)
  - Satellite images stored instead
- **Container**
  - Docker image for ISD & XNAT
    - ISD : commercial package for medical image annotation (IntelliSpace)
    - XNAT : open source counter part of PACS, medical database





UCSD updates

PRAGMA Cloud Scheduler GUI

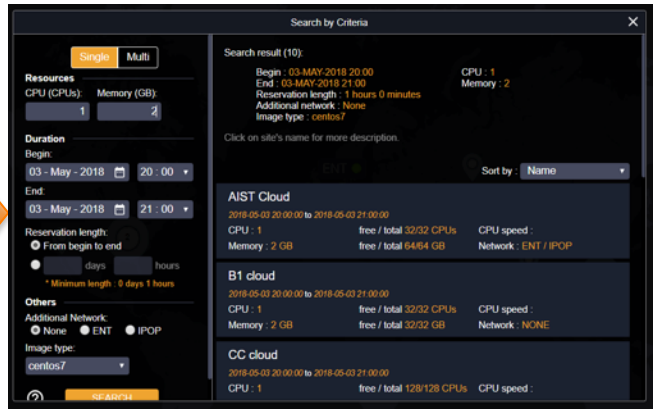
- **Admin interface**

  (student from Chungnam National University, Korea)
  - Manage users, **resources**, virtual images
  - Different views of cloud testbed:
    - User: what cluster is running and where
    - Admin: all virtual clusters and where they run
- **Multisite cluster reservation (see poster)**

  (students from Thammasat University, Thailand)
  - User interface: reserve a multisite cluster
  - Backend database extension
  - Algorithms to handle resources search and selection

Identity management

- Identity federation (EduGain), what is needed to join
- Tested app with KeyCloack an CILogon



GPU virtualization

- Run 2 virtual clusters each with 4 GPUs
- All software install is automated via rolls
  - Cuda : cuda toolkit and NVIDIA driver
  - Gpupt: GPU pass-through allocation
  - Tensorflow
  - Spark

Applications using Virtual GPU

- Traffic monitoring application to process movies from data sensors (see demo)
- Image classification analysis (see demo)

Kubernetes

open source system for automating deployment, scaling and, management of containers

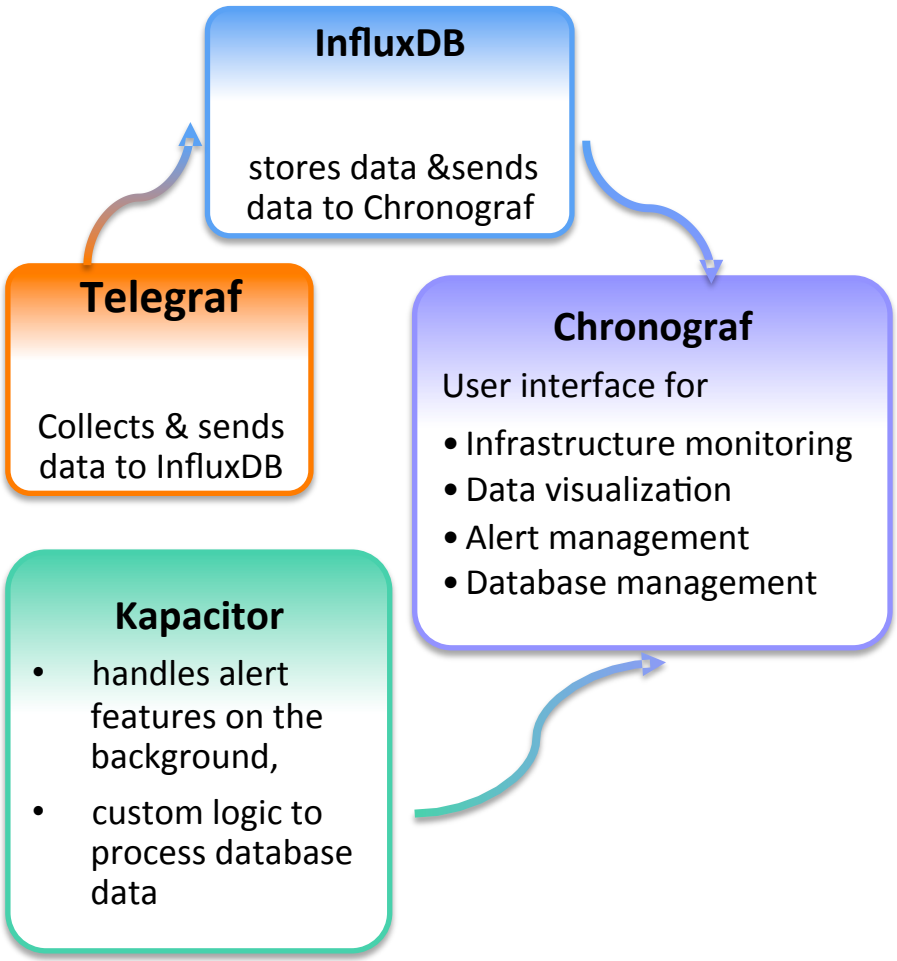
Kubernetes roll
- install and configure kubernetes on a virtual cluster.
- Configure local docker repository
- Add calico network for kubernetes pods

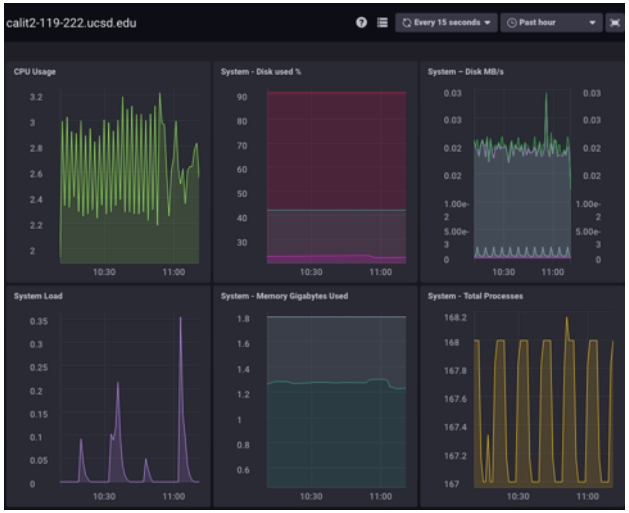
Run Condor pool (multiple containers) inside kubernetes cluster (see demo)

UCSD updates: Tstat

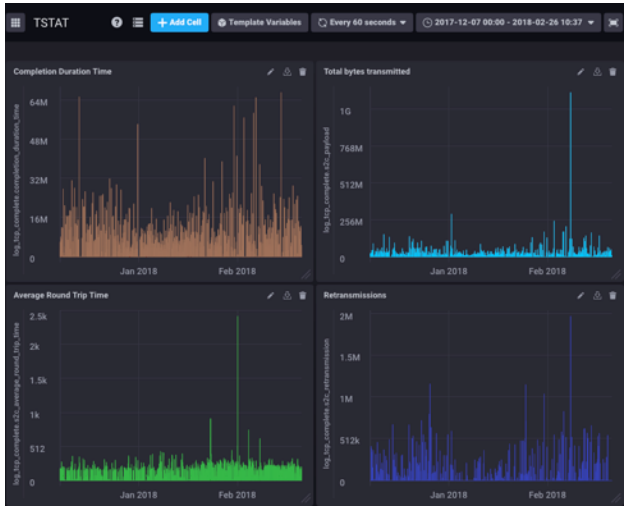
Tstat – TCP statistic and analysis tool http://tstat.polito.it/web.shtml

- roll https://github.com/pragmagrid/tstat
- Visualization with InfluxData TICK stack
- UCSD-UFL data transfer and tstat collection





Cpu usage
Disk used
System load
Memory used
Total processes



Completion time
Total bytes
System load
Round Trip Time
Retransmissions

UFL: Virtual GPU

- Ran neural network model on a virtual GPU cluster to classify images using Spark to distribute work to multiple nodes
- Saw a 10x speedup of model training with GPUs vs with CPU only (see demo)

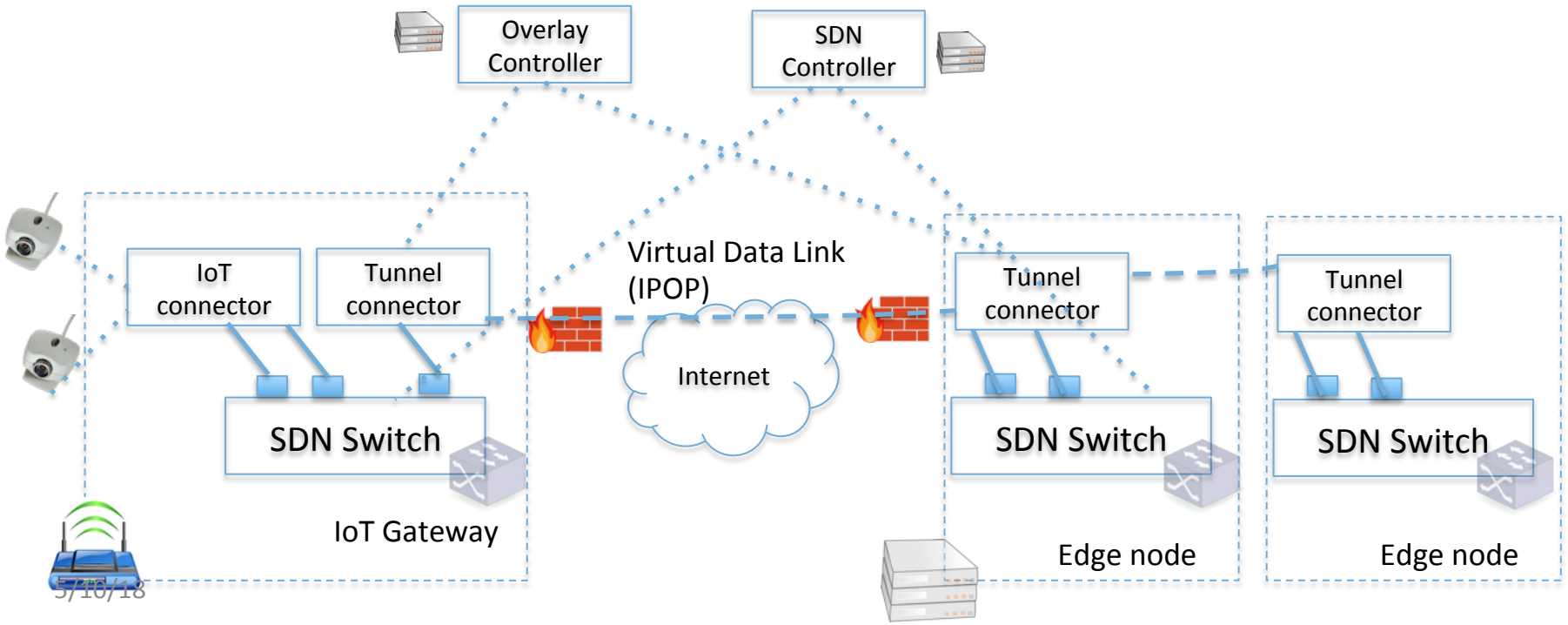
Applications of deep convolutional neural networks to digitized natural history collections

Eric Schuettpelz,1 Paul B. Frandsen,2 Rebecca B. Dikow,2 Abel Brown,3 Sylvia Orli,1 Melinda Peters,1 Adam Metallo,2Vicki A. Funk,1 and Laurence J. Dorr1

UFL: IPOP updates

Major effort towards Spring'18 IPOP release

- Support to link IPOP endpoints to virtual switch ports
  - Open vSwitch, Linux bridge
  - Every end of an IPOP link terminates in a tap device than is dynamically plugged to the switch port
  - Switch itself can be programmed using SDN



Lake Expedition Updates

- GRAPLEr software, resources, uses
  - Usability improvements, bug fixes
  - Research with land use & climate change scenarios
  - New GRAPLEr teaching modules being developed (led by Cayelan Carey & Kait Farrell)
    - *NOW* being taught on the Pacific Rim! Griffiths University (Australia), Washington State University, + 11 others
- Extending towards the edge
  - NSF S&CC water quality forecast
  - IPOP sensor gateway
- Science applications
  - See Kait Farrell's poster!

Biodiversity Expedition Updates

Collaboration

- Lifemapper installation at NCHC and populated Taiwan-focused data
- Presented a workshop at NCHC for local researchers based on local installation

Infrastructure

- Updated physical host cluster for virtual clusters to Rocks 7.0
- Installed, populated, computing North American data on XSEDE Comet cluster

Lifemapper code

- Deployed new code base to production
- Added Meta-community Phylogenetic Analysis (MCPA) tools
- Optimizing code and configuration for different size data and resources

User Interface (see demo)

- New browser-based interface, more accessible for new users
- Focused on enabling for SAGE2, working on assembling SAGE2 wall components
- Developing training materials

Next Steps

- Create API-interfaces linking Lifemapper with iDigBio and OpenTree for live user queries
- Generalize Comet configuration/installation for new instances / different scales