

# **PRAGMA 35**

## **RESOURCES & DATA WG**

## **AND EXPEDITIONS**

## **UPDATES**

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

# Resources WG: what we planned during PRAGMA 34

1. Data
2. Kubernetes , containers, VMs
3. EduGain/EduRoam
4. S3
5. IPOP

# Data transfer UCSD-ABCI

- 60Tb, each chunk 30Tb
    - 1 complete experiment after processing is 30 Tb (~ 70k time series)
    - Ideally to move entire experiment data set
  - Work assignment: (Salk/AIST/UCSD)
    - Preprocessing in SD
    - Move data from USB drive to UCSD
    - Transfer from UCSD to ABCI
      - Use Fast Data Transfer (FDT)
      - ABCI gateway node
      - Tstat to collect logs
- Gateway node:**
- GPFS volume is mounted
  - 5Tb can be used from the gateway / compute nodes
  - CPU: Intel Xeon E5-2640 v4 (2.4GHz, 10 cores) x 2
  - Memory: 256GiB
  - HDD: SAS 300GB
  - Network: InfiniBand-EDR x2, 10GBASE-T x2, 1000Base-T x2

# Japan-Taiwan Data and AI module platform for Analyzing Remote Sensing Data

Hidemoto Nakada, Ryosuke Nakamura, Ryousei Takano, Yoshio Tanaka (AIST)  
Hsi-En Yu, Chun Hung Huang, Chia-Chuan Chuang, Zhou-Jin Wu, Weicheng Huang  
(NCHC)  
Bo Chen, Scarlet Peng (NSPO)

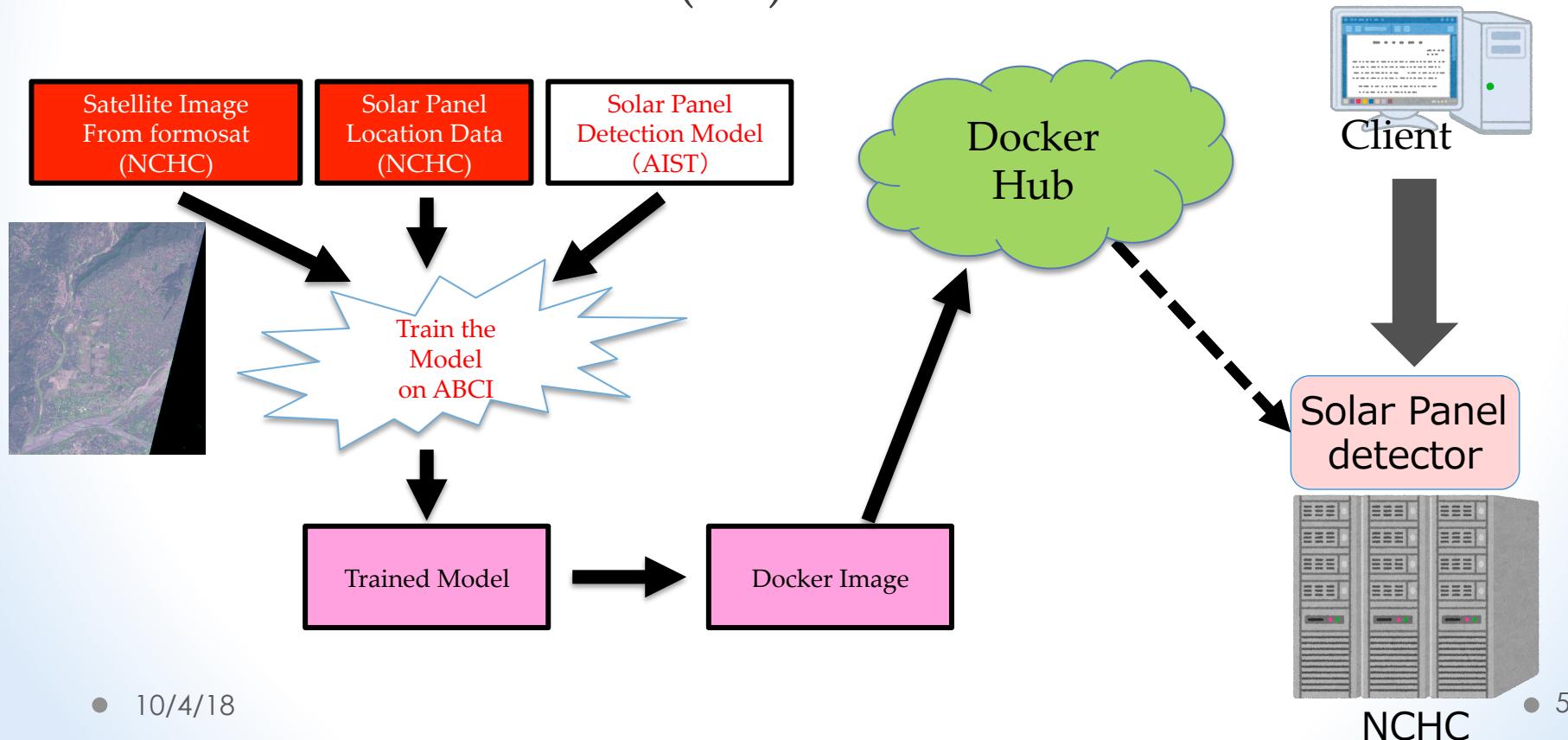
- AIST AIRC publishes / maintains AI modules as Docker images
  - Ex. Solar panel detection from satellite image.
- Confirm the validity of Docker Image as publication media.
- Confirm that the model developed for Japanese solar panels is valid for Taiwanese solar panels

This work is based on results obtained from a project commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

# Demonstration Configuration

Demo during Resources WG breakout, Oct 4

- Deploy the preconfigured 'Module' docker container on NCHC node
- Use it from the client (PC) here.



# Kubernetes, Containers, VMs

- Jupyter-based SDK is runnable on a VM or a container:
  - Controls the deployment of international L2VLAN and Inter-Cloud environment (NAIST)
- Build k8s as a virtual cluster and move anywhere with `pragma_boot` (UCSD)
  - Have k8s cluster
  - Finalizing DNS k8s container setup and implementation
- Container with JupyterLab for using Deep Learning for 2018 APAC HPC-AI Student Competition (UCSD/Thammasat University)
  - Tensorflow + cuDNN
  - GPU
  - CILogon for authentication
- Implementing Indonesian e-health community cloud model on PRAGMA cloud (UCSD/Universitas Yarsi, Indonesia):
  - Run 3 “attacker” VMs
  - Run 2 “server” VMs
- Work in progress: add container support to `pragma_boot` (UCSD)

# EduRoam, EduGain, CILogon

- If an Institution is a member of EduRoam then can use EduGain
- PRAGMA members should become members of EduGain.
  - Number of requirements
  - Facilitates all logins and authentication
  - Contact with Jim Basney
  - Make a list of steps (logical) and follow up in email and also add to PRAGMA website (TODO)
- Can use CILogon as an authentication mediator  
Example: JupyterLab container (UCSD)
  - Register client with CILogon delegation service <http://www.cilogon.org/oidc>  
Provide
    - Client name PRAGMA AI Jupyterhub
    - Contact email nwilliams@ucsd.edu
    - Home URL <https://pragmagpu.nautilus.optiputer.net>
    - Callback URL [https://pragmagpu.nautilus.optiputer.net/hub/oauth\\_callback](https://pragmagpu.nautilus.optiputer.net/hub/oauth_callback)
  - Once approved will receive
    - Here is your client Identifier: [myproxy:oa4mp,2012:/client\\_id/22a90ed...b1e022911cef4](myproxy:oa4mp,2012:/client_id/22a90ed...b1e022911cef4)
    - Here is your client secret: [t7PeYnJN-usF...qcljofQR4T\\_W6ShzHa9r9kQO6q89XfyA](t7PeYnJN-usF...qcljofQR4T_W6ShzHa9r9kQO6q89XfyA)
  - base64 encode id, secret and callback

# S3

- Work on issues in connection of container to s3 storage (AIST)
  - Can mount s3 bucket outside of the container, not inside the container
    - For the fuse mount issue in the container, we need to upgrade Linux kernel to the version 4.18 or higher.
    - However, not confirmed yet if it has been really solved in the current environment.
- User authentication
  - try KeyCloack <https://www.keycloak.org> and CILogon <http://www.cilogon.org/>
  - AIST: happy to continue the work because the SSO function is necessary to promote S3 use in multiple sites.
  - UCSD: CILogon authentication with JupyterLab container

# Work with undergraduate students

- Students
  - Thammasat University (Thailand) / UCSD / AIST
  - CNU (Korea) / UCSD
- Projects
  - Data storage monitoring and visualization
  - Applications performance monitoring and visualization
  - PRAGMA Cloud scheduler administrative interface

# IPOP v18 release

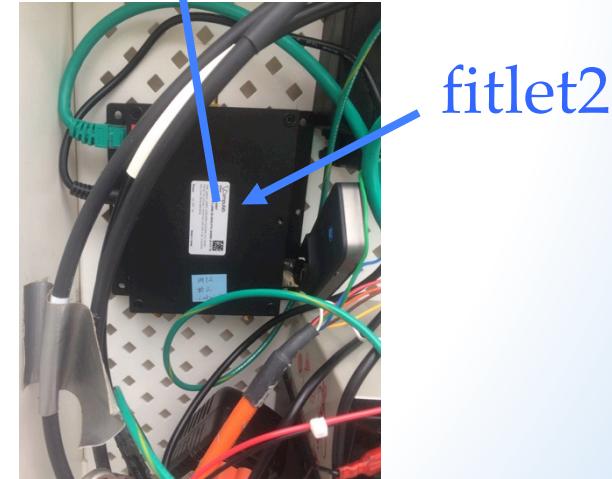
- Switching: SDN (open vSwitch)
- Rather than multiplexing a tap, dynamically manage taps connected to SDN switch ports
- Layer-2 switching; tested with OVS and STP
- Significant development/testing effort
  - See Ken's demo

# Expeditions Updates

Lake Ecology  
PRAGMA-ENT  
Virtual Biodiversity

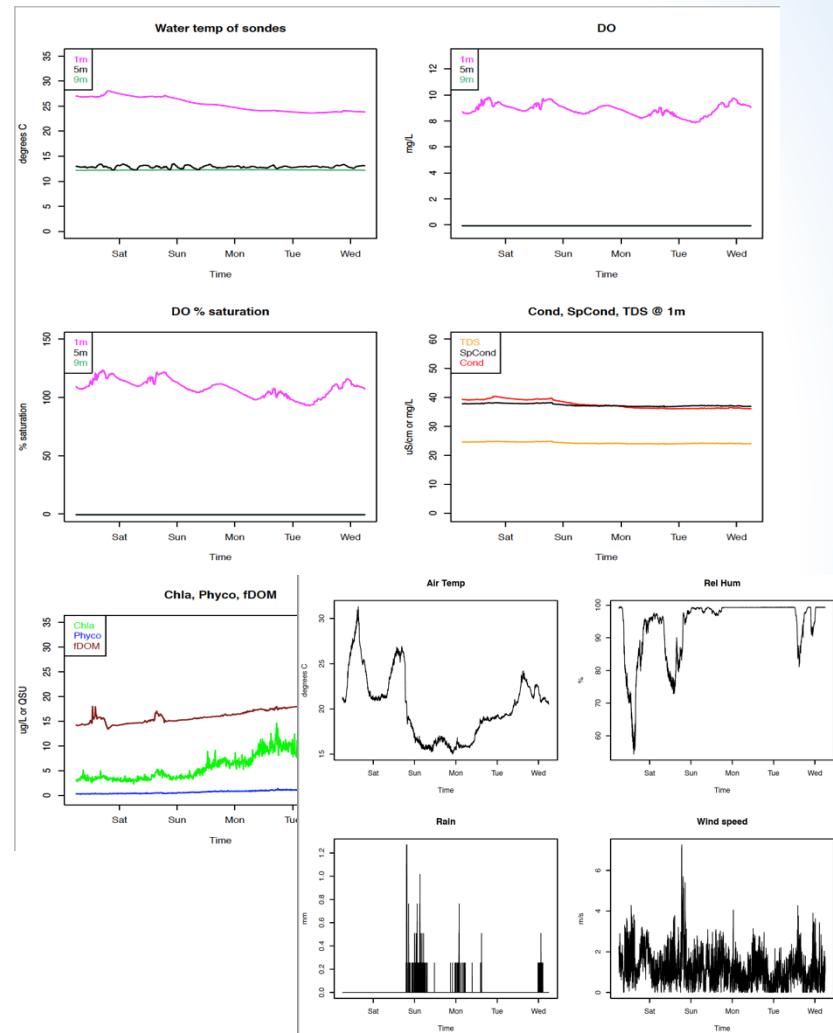
# Lake Ecology: Sensor Gateway Design

- Fanless mini-PC, open-source software
- Reads/stores data from logger (Campbell)
- Security, NAT traversal:
  - Connects to IPOP overlay VPN via 4G cellular link
  - Data publishing, remote maintenance
- Pushes data updates (diffs) to private and/or public repository using git



# Prediction Workflow

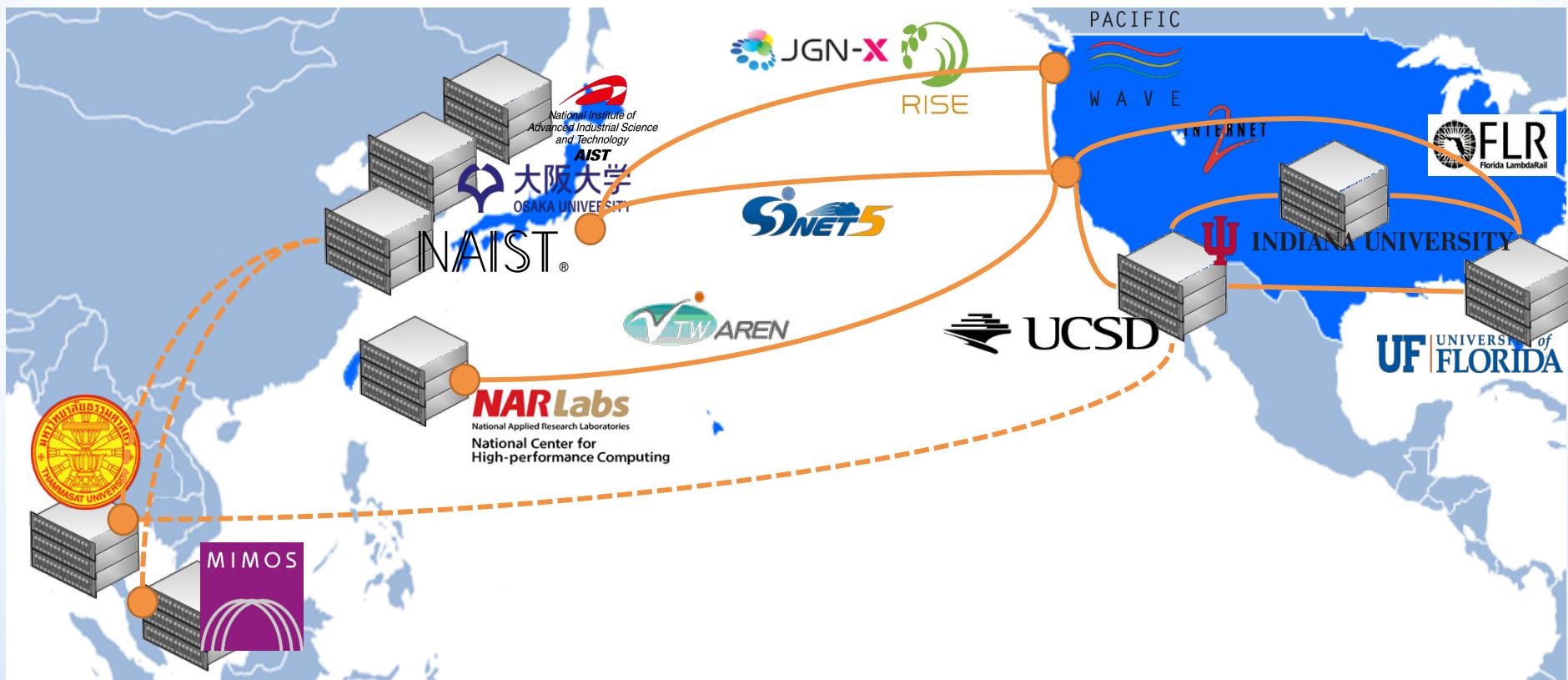
- Data collected from multiple sensors (air temperature, rain, wind, water temperature, DO, chlorophyll) committed to git data repos over IPOP daily
- Triggers pull execution of ensemble GLM forecast model runs
  - Dedicated servers, or GRAPLER



# PRAGMA-ENT updates

## Infrastructure

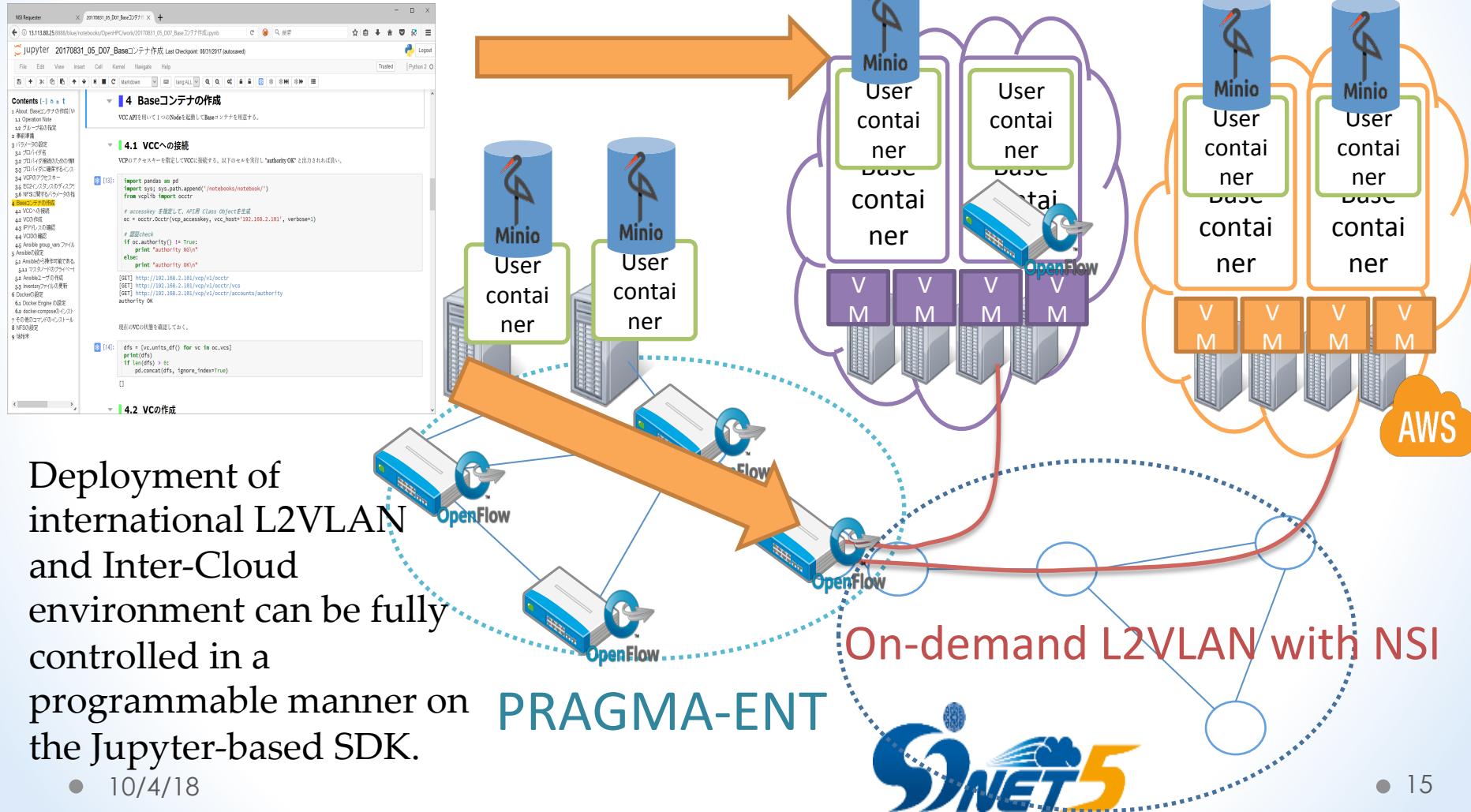
- SINET link (another Japanese NREN) can be dynamically controlled and deployed in a programmable manner.
- fully programmable interface for integrating ENT and SINET infrastructure



# Inter-Cloud environment deployment using dynamic L2VLAN service

## Jupyter-based SDK

NAIST: Demo Oct 4



Deployment of international L2VLAN and Inter-Cloud environment can be fully controlled in a programmable manner on the Jupyter-based SDK.

# PRAGMA-ENT



# VBE updates

Hsiu-Mei Chou, NCHC  
Aimee Stewart, KU  
Nadya Williams, UCSD

KU: more info during  
WG breakouts

- **KU - UCSD July 2018 meeting outcomes**
  - Dynamic load balancing and job submission optimization
    - Aggregate elements of processing by species for less fine-grained monitoring and resource utilization but simpler workflows with less IO
    - Move long-running frontend processes to compute nodes
    - Write outputs from compute nodes directly to final, shared data space
  - Identify resource requirements for dataset footprint
    - SGE / Makeflow tools to aggregate usage stats
  - Miscellaneous troubleshooting
    - Create different configuration for workshop vs production installation
    - Identified some tools troubleshooting inconsistent failures

# VBE updates (cont'd)

- **KU - Lifemapper progress**
  - Focus on production level reliability
    - Handle large data inputs
    - Handle 3<sup>rd</sup> party software failures
    - Harden workflow recovery
  - Simplify data ingestion
    - Fewer requirements for species data
    - Use the same processes for data ingestion for public data and user data
  - Improved User Interface
    - Simple form for user data ingestion
    - 3-way linked visualization – map, phylogenetic tree, scatterplot (in progress)
    - Speed up map displays
- **NCHC – UCSD– KU work**
  - National Museum of Marine Science and Technology
    - “Exploring coastal species of Taiwan” funded September 2018
    - Meeting in late 2018/ early 2019
    - New Lifemapper virtual cluster to be installed within 6 months
    - Workshop spring 2019