

Resources Working Group Updates

Philip Papadopoulos (UCSD)
Yoshio Tanaka (AIST)
Co-Chairs



PRAGMA 32
Gainesville, Florida

Update Topics

Cloud Scheduler – Demo this Morning

PRAGMA ENT PRAGMA

Data Server @ UCSD (160TB raw/120TB Usable)

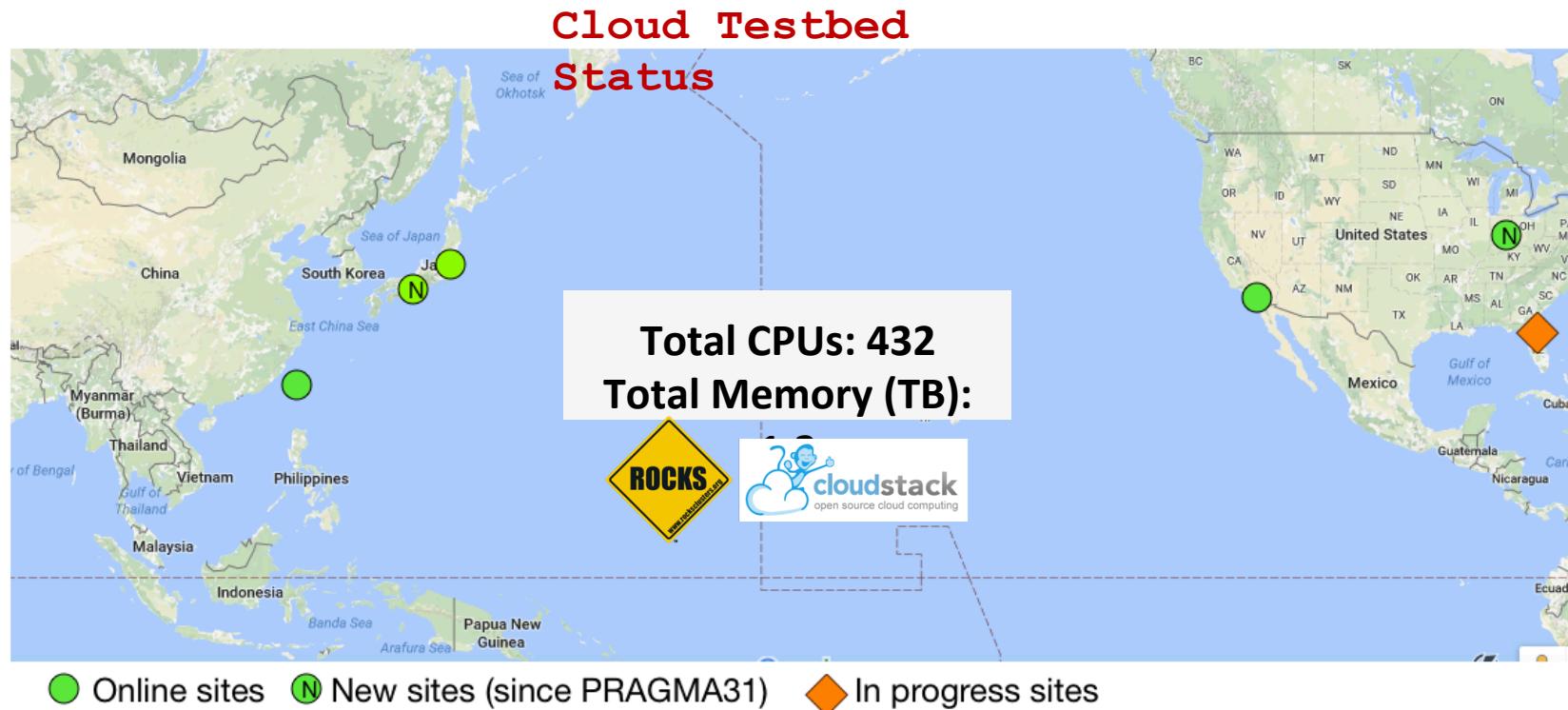
LifeMapper – Comet Virtual Clusters + Code Updates (Biodiversity)

iPOP and GRAPLER Updates (Lake Ecology)

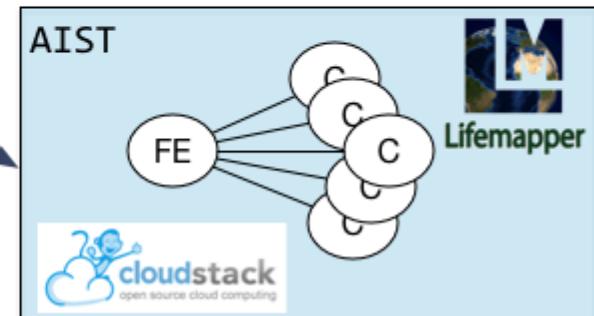
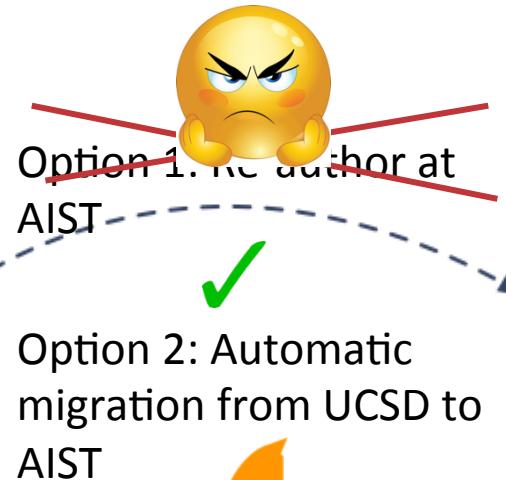
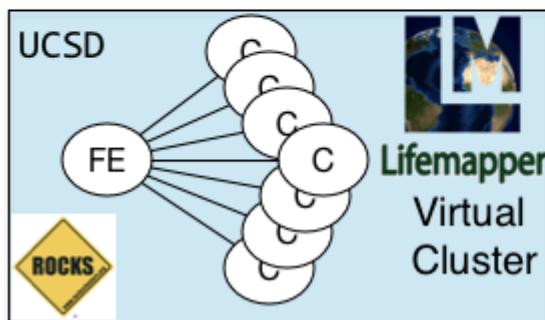
Working Group “Agenda”

PRAGMA Cloud Testbed

- **Goal:** A persistent Cloud testbed for Biosciences and other PRAGMA working group members to run application experiments.



Motivation



Format:
KVM + ZFS vol

Properties:

FQDN
IP Address
Assignment
Number of compute nodes

Universal Format

PRAGYA
BOOT...
Localize configuration

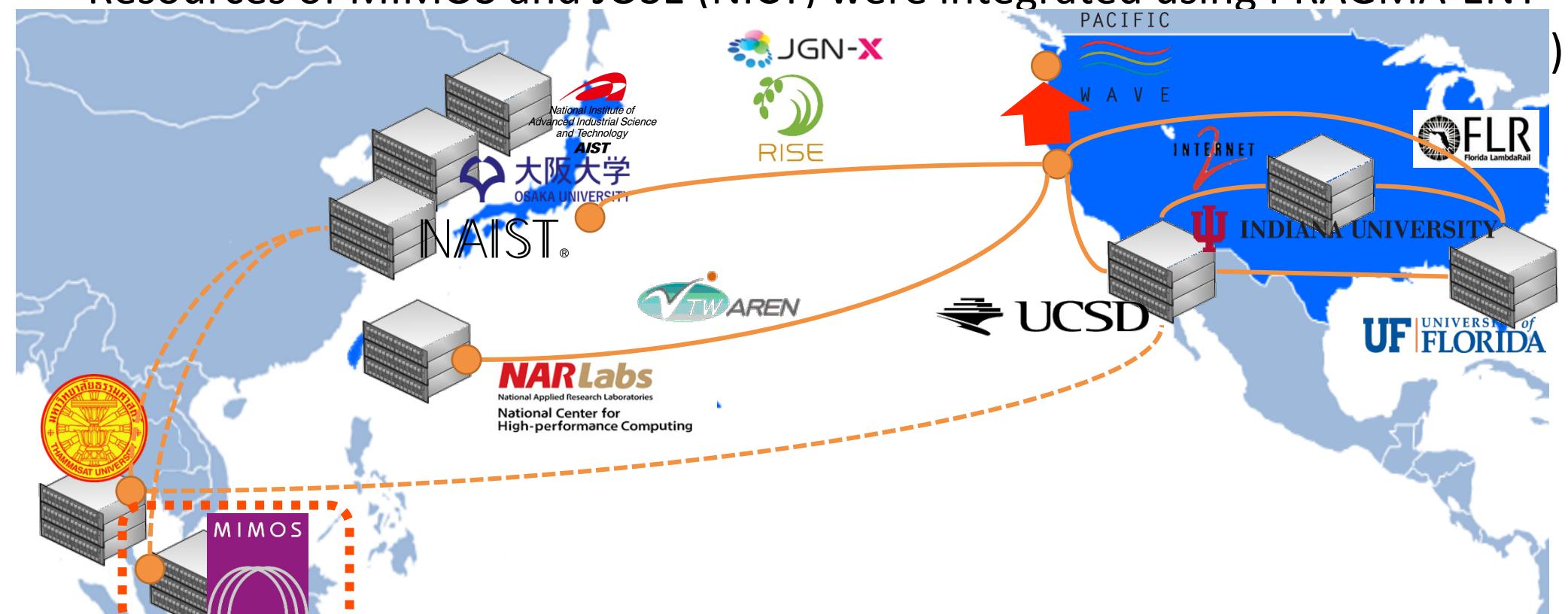
Format:
KVM + OCOW2
Properties:
FQDN
IP Address
Assignment
Number of compute nodes

When (and where else) can I run my virtual cluster? cloud scheduler.

Updates of PRAGMA-ENT

Infrastructure

- ENT backbone was extended to MIMOS, Malaysia with GRE
- Resources of MIMOS and JOSE (NICT) were integrated using PRAGMA-ENT



Updates of PRAGMA-ENT

Applications

- Software Defined Storage
 - Luke (MIMOS) has been deploying Software Defined Storage System using JOSE (NICT's virtual computing platform) with PRAGMA-ENT
- Remote visualization for disaster management
 - Watashiba (NAIST) has been deploying SAGE2 visualization environment over PRAGMA-ENT
- Low overhead network virtualization in multi-tenant Cloud data center
 - Kyuho (UF) will give a demonstration [[Demo on Fri.](#)]

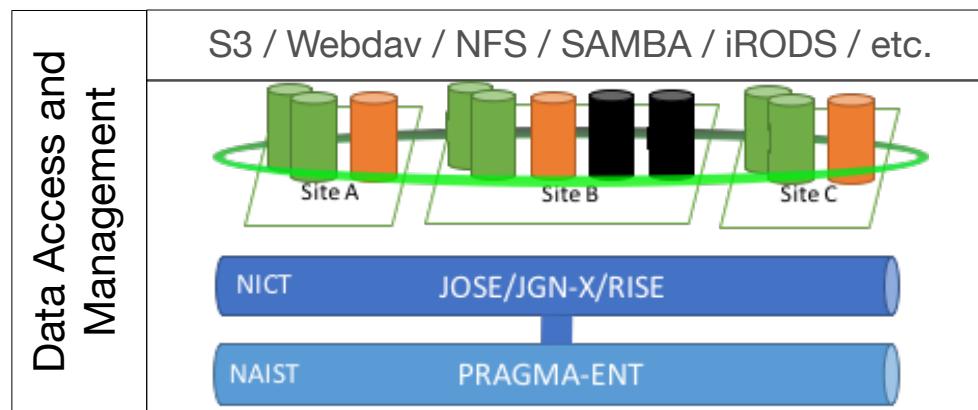
Data Visualization & Software Defined Storage

Visualization

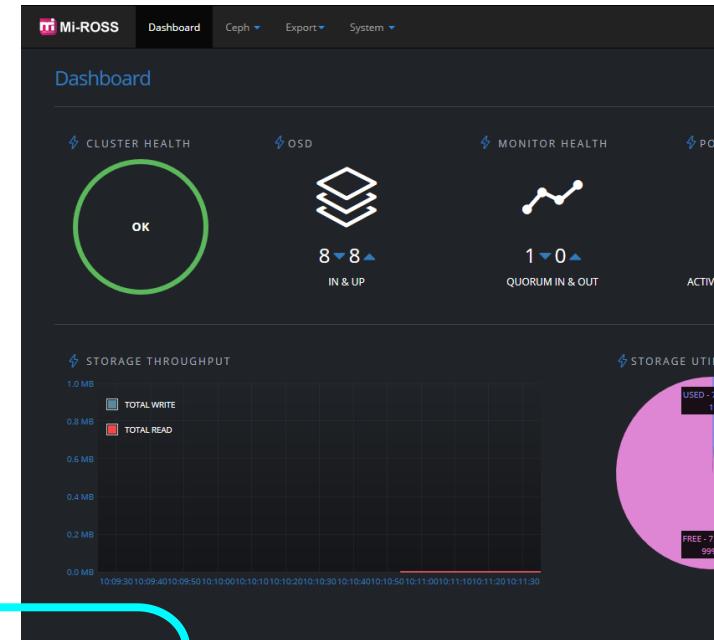


SAGE2

Based
on
the
Distributed
Object Storage
System
(DOSS)



More apps: e.g. 3Di Water Management, IFAS (Integrated Flood Analysis System), NOAH, etc.



Source(s)



A S T I

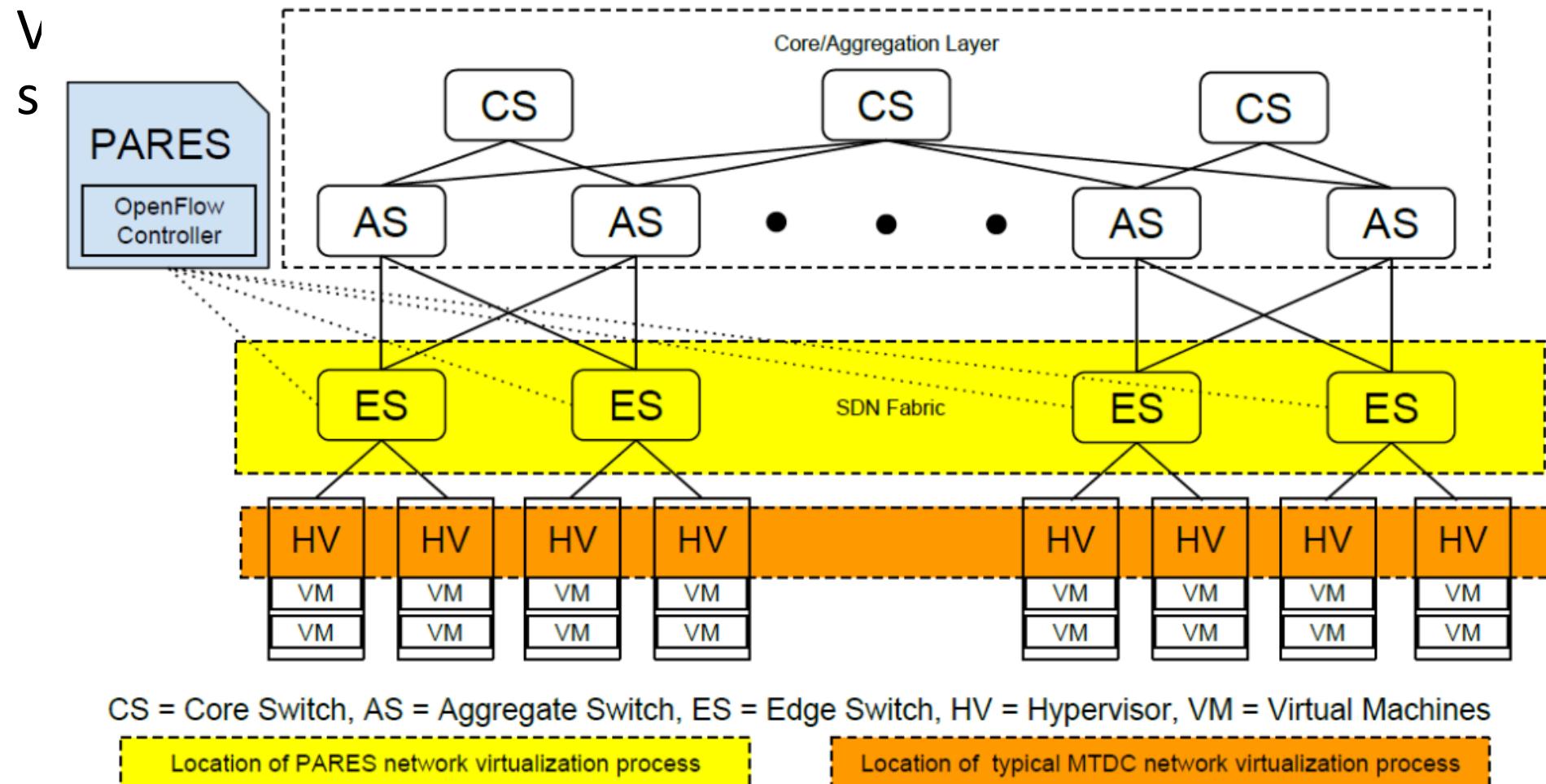
NOAH Data



Dam Data /
Weather Data

Others: AirBox,
River/Flood
Models, etc.

PARES: Low overhead network virtualization in multi-tenant Cloud data center [Demo on Fri.]



PRAGMA Data Server @ UCSD

Based upon FIONA (Flash-IO Network Appliance) built in the Pacific Research Platform
120 TB usable storage, ~2GB/sec.



16 X



1Socket E5 1620-v3

Lifemapper PRAGMA 32 Update

Deploying Lifemapper on XSEDE resource Comet

- Address a large number of small to medium size computations
- Access large data storage over 10Gb network



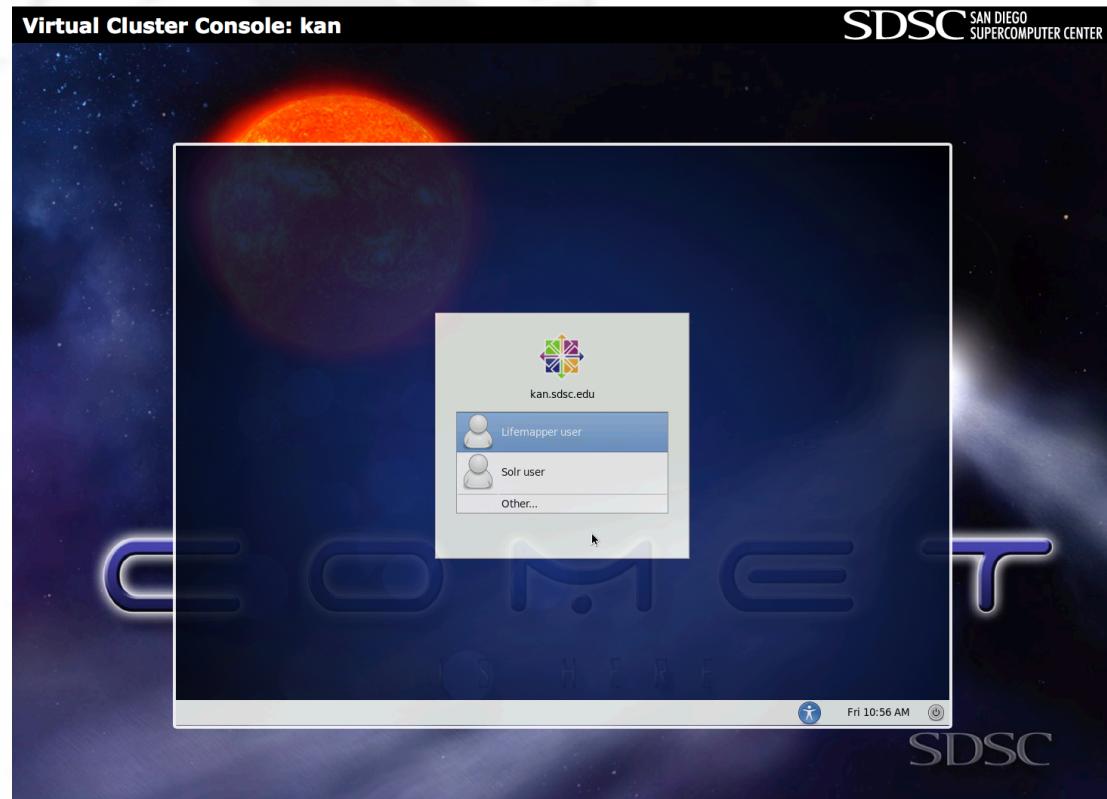
Comet Deployment

- Move to Comet to address large projects
- 12,000 species x 15 climates x 3 environmental sets
 - $(12,000 + (15 * 12,000)) * 3 = 576,000$ computations
- Allows us to deploy Lifemapper as an integrated whole
- CC Tools for job management
 - Makeflow
 - Work Queue
- Simple virtual cluster Request-Deploy-Retrieve

Comet Virtual Cluster

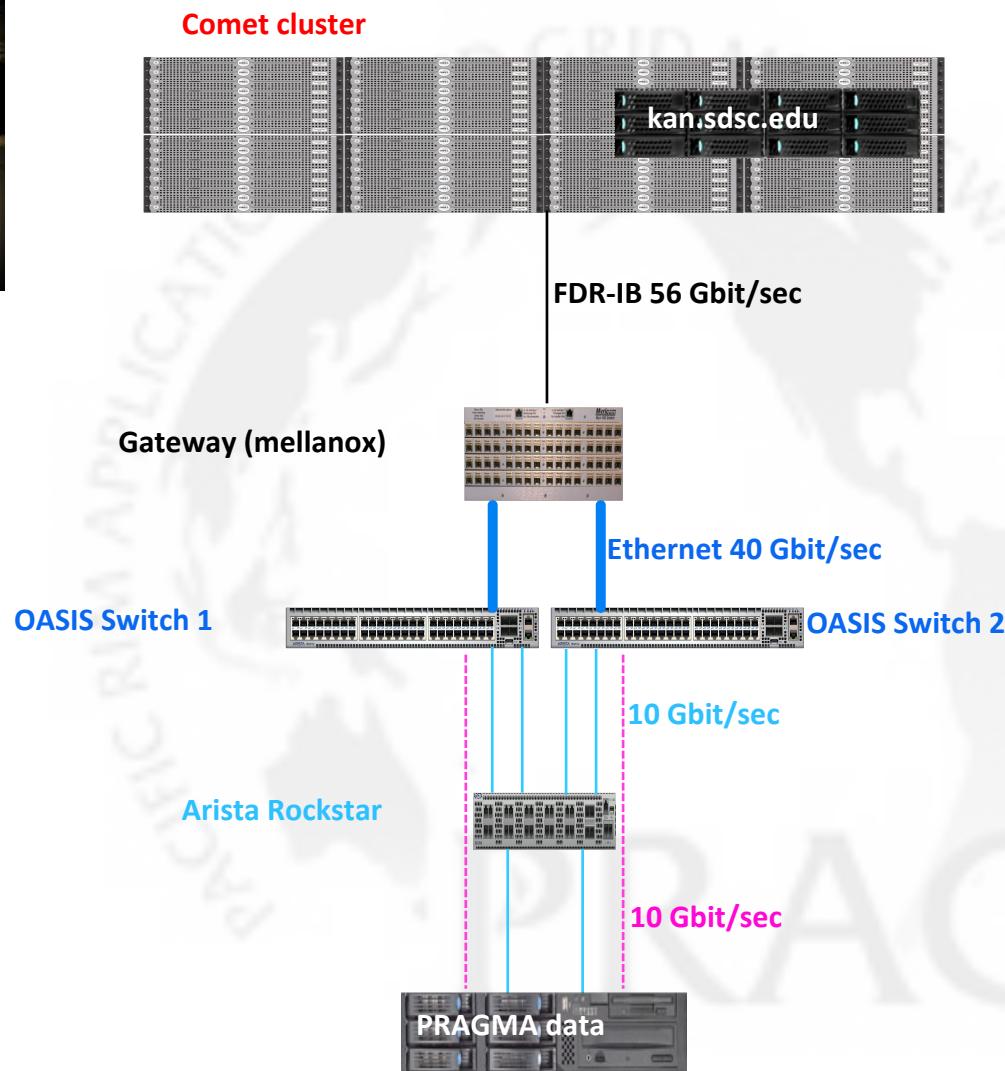
- Virtual machines

Name	Kind	CPUs	RAM(Gb)	Disk(Gb)
Kan	1 Frontend	4 E5-2640 v3 @ 2.60GHz	32	36
Vm-kan-0-x	8 compute	24 E5-2640 v3 @ 2.60GHz	96	36



- Software stack
 - Rocks 6.2
 - SGE
 - Web server
 - Lifemapper Server
 - Lifemapper Compute
- See cluster install movie
<http://goo.gl/fPiYtj>

Virtual Cluster Network



- PRAGMA data 120Tb (raw 160Tb)
- NFS read/write ~ 5 Gbit/sec from kan.sdsc.edu to PRAGMA data set
- Iperf
 - 1 stream ~ 9.8 Gbit/sec
 - 2 streams ~ 15.5 Gbit/sec

Lifemapper Code Update

- Bug fixes
- Finalizing indexing with **solr**
- Continue with enabling Lifemapper configuration and all setup/test invocations as a command line infrastructure. This will simplify all commands and keep install, Configuration, testing unified and extensible.
Example:

```
lm init db  
lm start/stop pipeline  
lm list users  
lm update ip
```

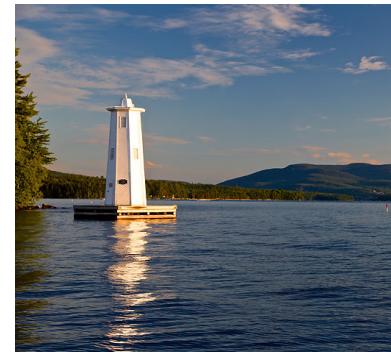
- Continue integration of **CC tools** for job management:
 - Makeflow
 - Work Queue
- Continue work on formalizing requirements and code for fully described data allowing easy use of different input datasets

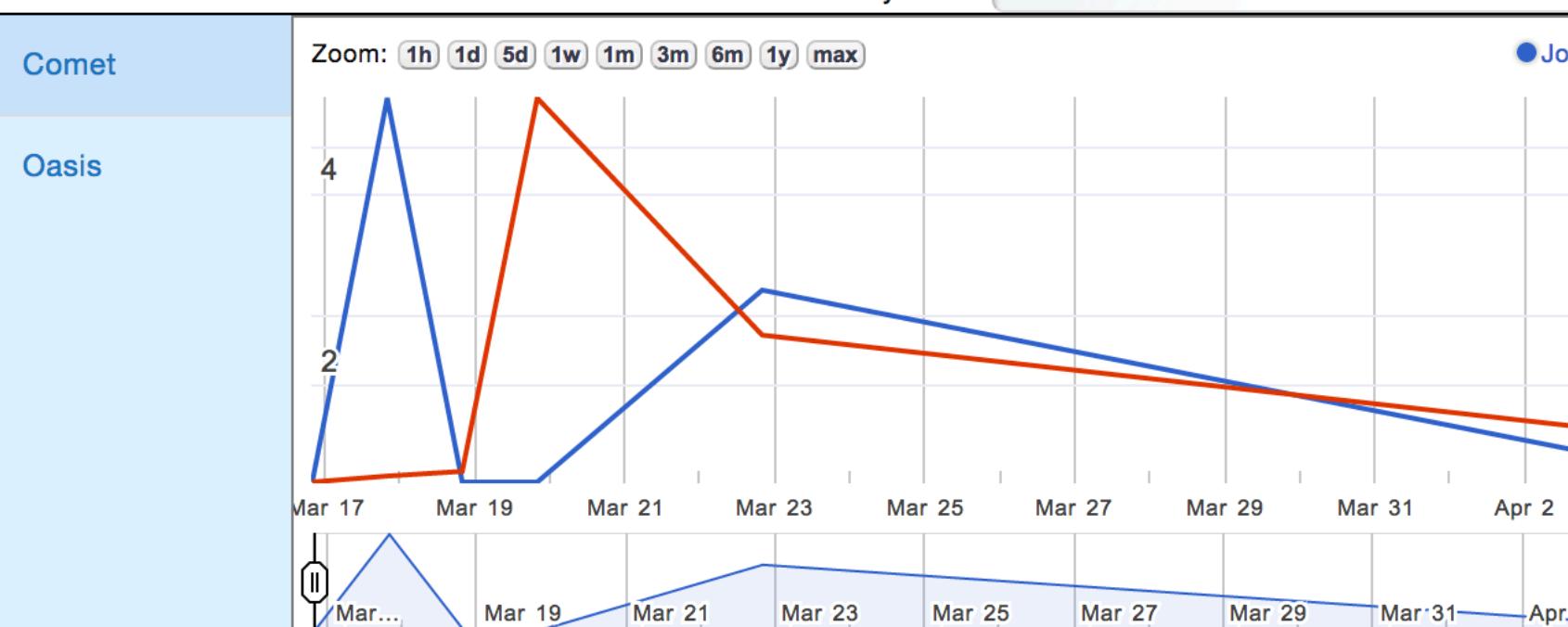
Updates on IPOP and GRAPLER resources

Expanded GRAPLER resources to use Comet/SDSC as an on-demand virtual cluster

- Worked through initial setup and transition to production Comet
- Virtual machines at UF provide Web service front-end and baseline pool of HTCondor resources (48 cores) for GRAPLER
- Comet nodes connect/add to the pool through IPOP
 - E.g., during lake expedition pre-workshop preparations – 100k+ GLM model runs

Maintenance and improvements for usability, performance of GRAPLER service





Total Owner Claimed Unclaimed Matched Preempting Backfill

/LINUX 48 0 0 48 0 0 0

Total 48 0 0 48 0 0 0

Updates on IPOP and GRAPLER resources

IPOP code updates

- Syncing with latest WebRTC open-source libraries
- Changes to low-level packet capture/tunneling code (“tincan”) to better support layer-2 (Ethernet) virtual networking, and multicasting
- Goal: IPOP-enabled IoT edge devices (e.g. Raspberry Pi, Intel Joule) and gateways (e.g. OpenWRT routers) to create virtual private network overlays from sensors to edge to cloud

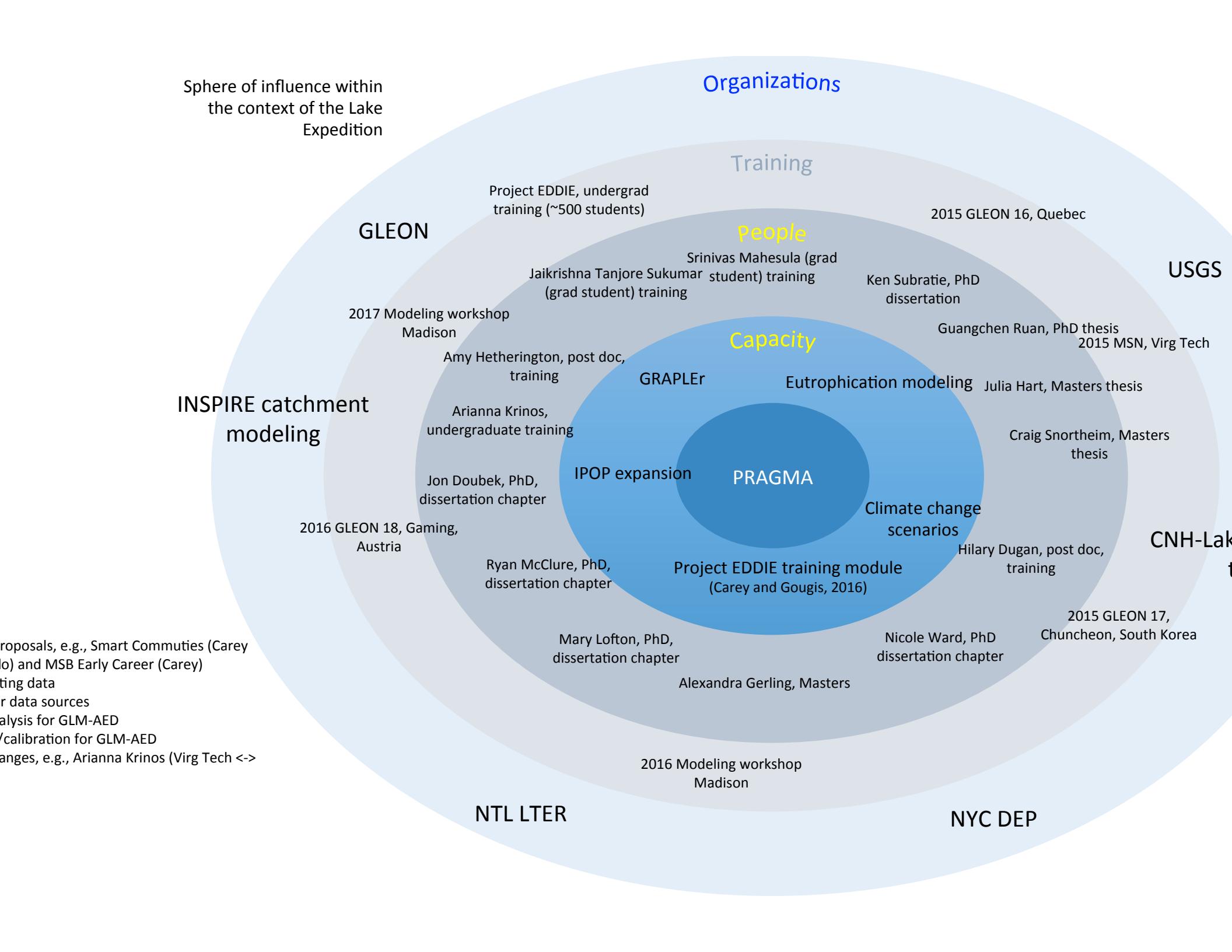
GRAPLER lake modeling efforts

100,000 simulations modeling the effects of climate change and land use scenarios on phytoplankton blooms in Lake Mendota (Cayelan Carey, Arianna Krinos, Amy Hetherington)

RAPLER lake comparison of nitrogen and phosphorus cycling in GLEON lakes under future climate scenarios (Cayelan Carey, Arianna Krinos, Karen Farrell, Nicole Ward, Mary Lofton, Ryan McClure, Jon Doubek)

Integration of GRAPLER into ecology classrooms in 5 universities in 2016-2017 academic year (several hundred students taught in total!)

Carey Lab grad students, 1 postdoc, and undergrad extraordinaire are submitting GRAPLER jobs regularly to advance lake modeling efforts



tie, K, Aditya S, Carey CC, Hanson PC, Figueiredo R. In
GRAPLER: A Distributed Collaborative Environment for
Ecosystem Modeling that Integrates Overlay Networks,
throughput Computing, and Web Services. Concurrency
computation: Practice and Experience.

Ecological Informatics 39 (2017) 10–22

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



J Sci Educ Technol (2017) 26:1–11
DOI 10.1007/s10956-016-9644-2

Recent publications by the Lake Expedition



Simulation Modeling of Lakes in Undergraduate and Graduate Classrooms Increases Comprehension of Climate Change Concepts and Experience with Computational Tools

Cayelan C. Carey¹ · Rebekka Darner Gougis²

Published online: 22 August 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Ecosystem modeling is a critically important tool for environmental scientists, yet is rarely taught in undergraduate and graduate classrooms. To address this gap, we developed a teaching module that exposes students to a suite of modeling skills and tools (including computer programming, numerical simulation modeling, and distributed computing) that students apply to study how lakes around the globe are experiencing the effects of climate change. In the module, students develop hypotheses about the effects of different climate scenarios on lakes and then test their hypotheses using hundreds of model simulations. We taught the module in a 4-hour workshop and found that participation in the module significantly increased both

Keywords Simulation modeling · Climate change education · Hypothesis-testing

Introduction

Motivation for EDDIE Lake Modeling Module

Environmental scientists are increasingly analyzing datasets of observations obtained through sensors and remote sensing, enabling new analyses and ecological phenomena (Hanson 2007; Porter et al. 2012). Conducting these anal-

Straw Topics for Resources Working Group

Integration of many “threads” of activity → More robust cloud infrastructure

- ENT
- IPOP
- Persistent Identifiers (PIDs)
- Network Measurement
- Cloud Scheduler
- → Application and infrastructure Testing

Discussion about adding distributed cloud storage capacity (~0.5PB) based on PRAGMA Data server + S3

From this morning: GPUs?

→ Set specific goals between now and Pragma33

→ Challenge to the “old” folks : Try to keep pace with the (very very impressive) students