Resources Working Group Updates

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

PRAGMA 32

Gainesville, Florida

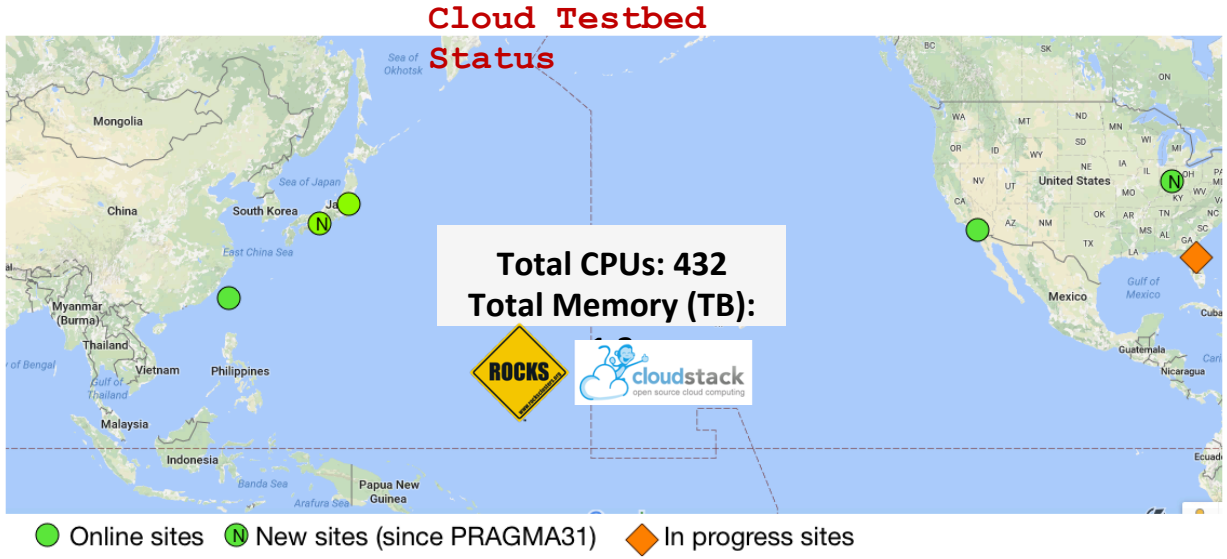
Update Topics

- 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)

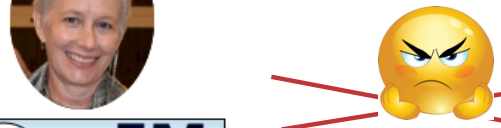
- Working Group “Agenda”

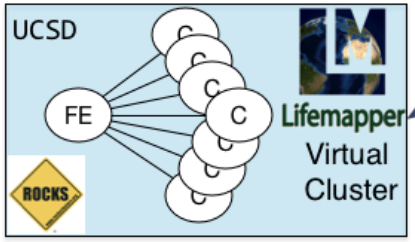
PRAGMA Cloud Testbed

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



Motivation

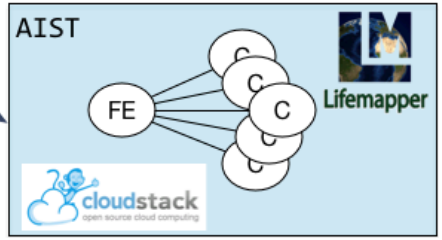




Option 1: Re-author at AIST

✓

Option 2: Automatic migration from UCSD to AIST



Format:
KVM + ZFS vol

Universal Format

Format:
KVM + QCOW2

Properties:
FQDN
IP Address
Assignment
Number of compute nodes

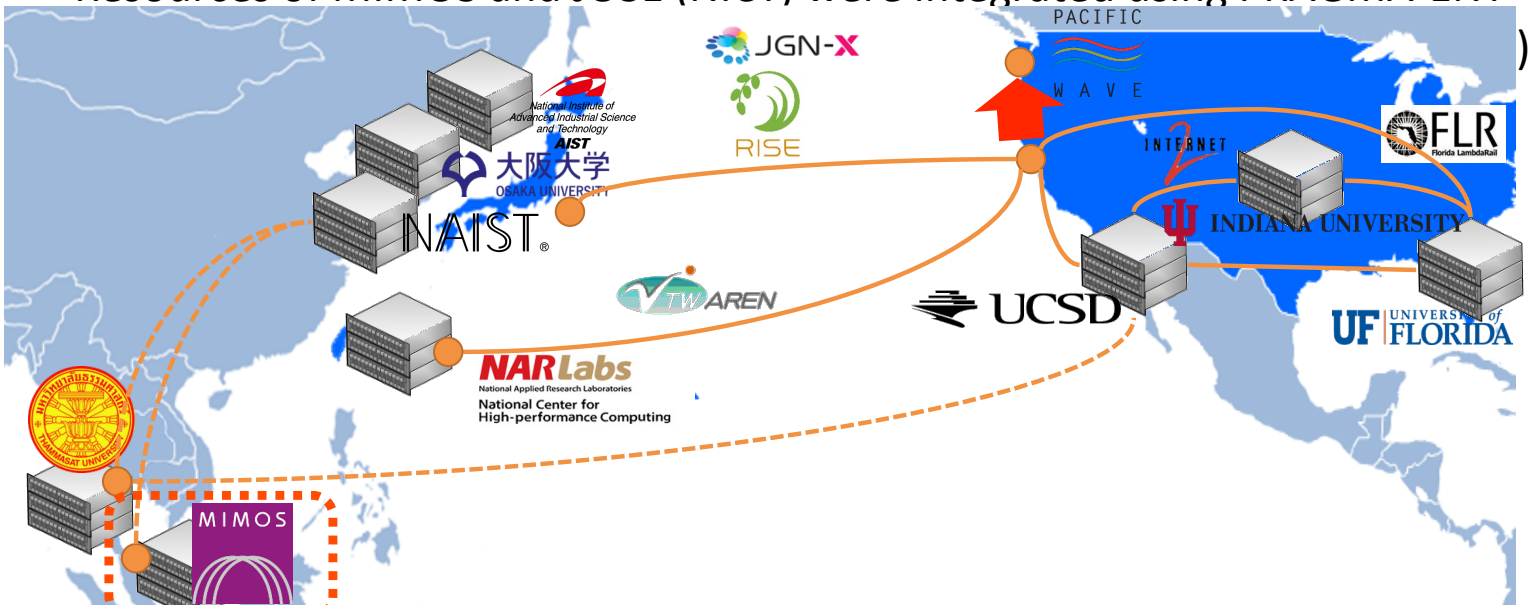
PRAGMA BOOT…
Localize configuration

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

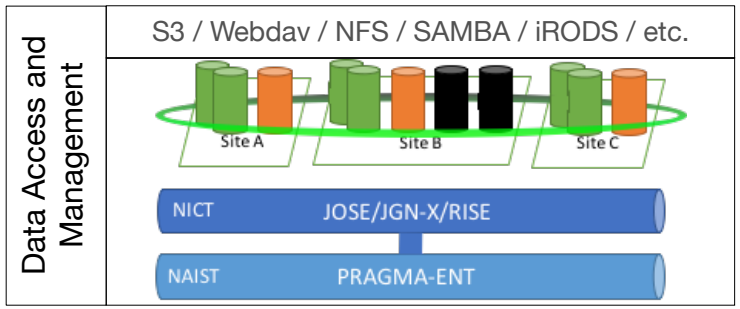
Data Visualization

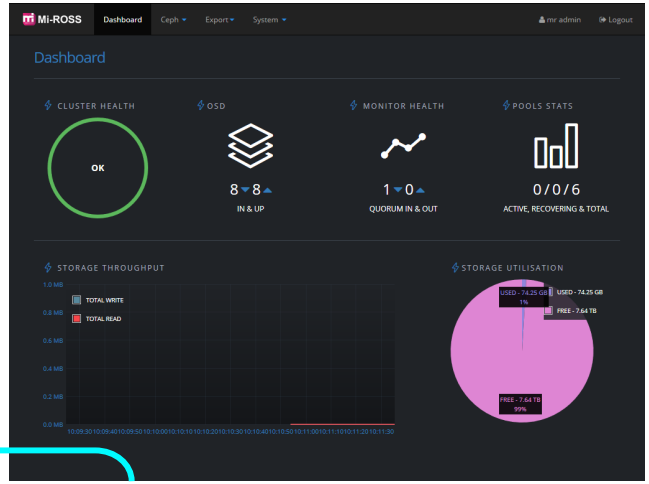
AIST
NATIONAL INSTITUTE OF ADVANCED INDUSTRIAL SCIENCE AND TECHNOLOGY (AIST)

SAGE2

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

WAN Based Reliable Distributed Object Storage System (Mi-ROSS)





Data Source(s)

ASTI

NOAH Data

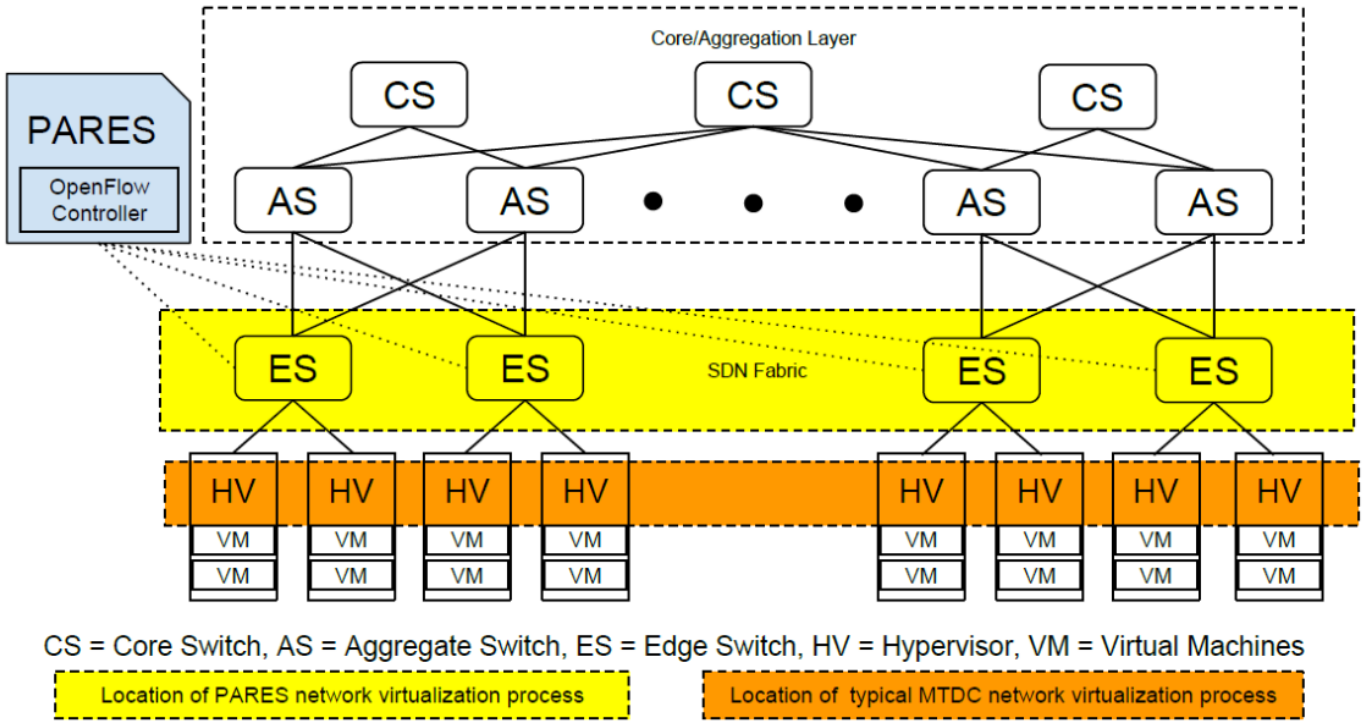
NECTEC
a member of NSTDA

Dam Data / Weather Data

Others: AirBox, River/Flood Models, etc.

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

- V
  s



PRAGMA Data Server @ UCSD

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

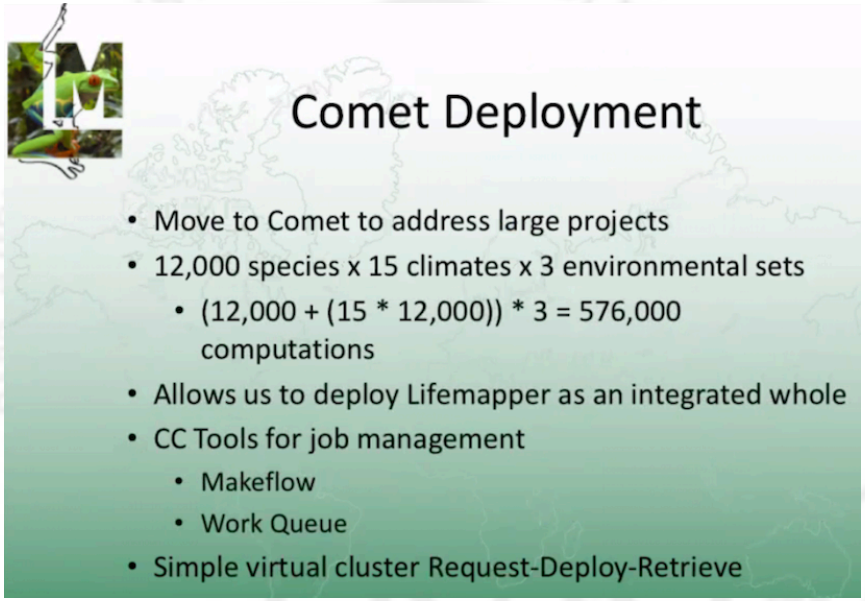
16 X

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

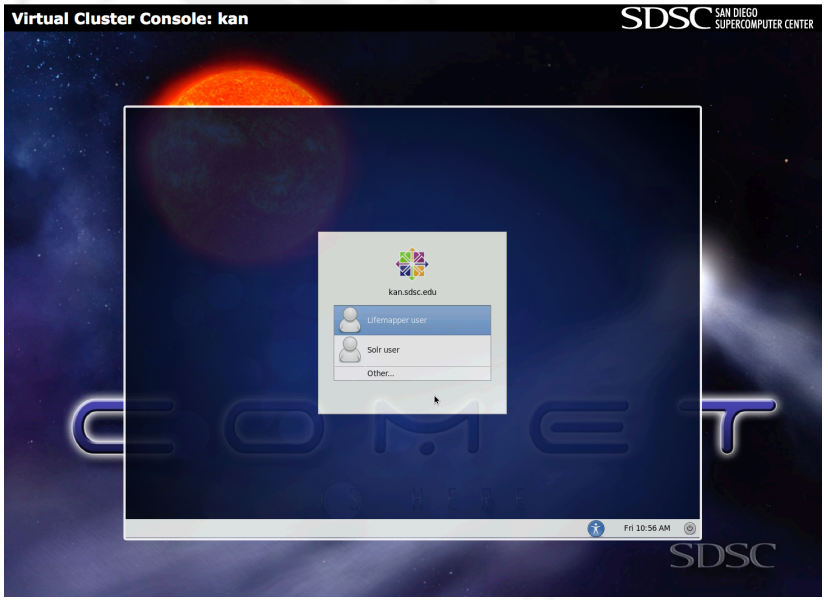


Comet Virtual Cluster

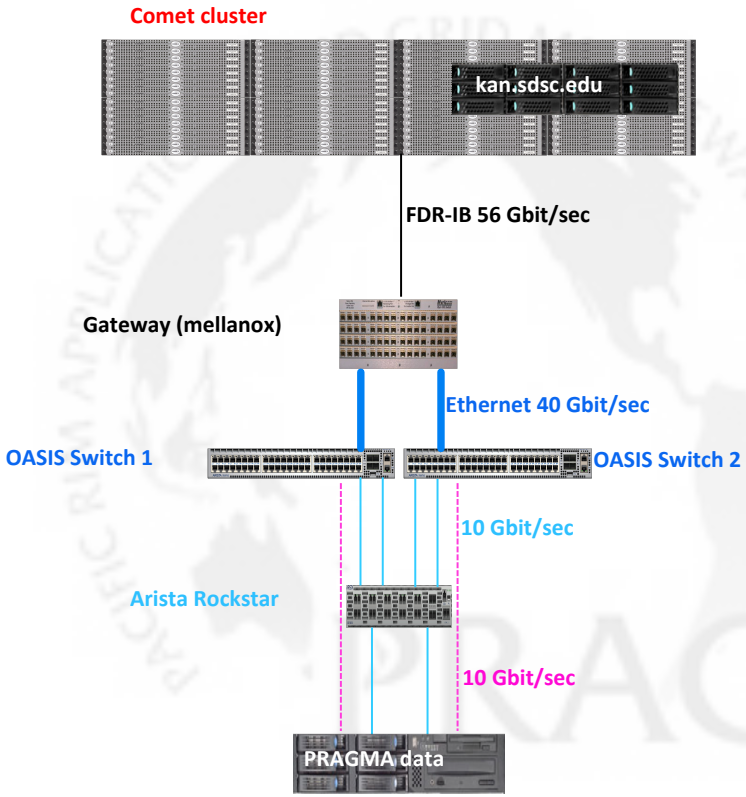
- Virtual machines

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) $10k
- NFS read/write ~ 5 Gbit/sec from kan.sdsc.edu to PRAGMA data server
- Iperf
  - 1 stream ~ 9.8 Gbit/sec
  - 2 streams ~ 15.5 Gbit/sec

Lifemapper Code Update

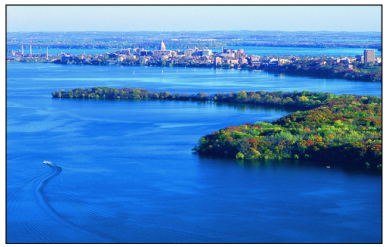
- Bug fixes
- Finalizing indexing with **solr**
- Continue with enabling Lifemapper configuration and all setup/test invocations as a command line infrastructure. This will simplify all commands and keep install. Configuration, testing unified and extensible. Example:

```
lm init db
lm star/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

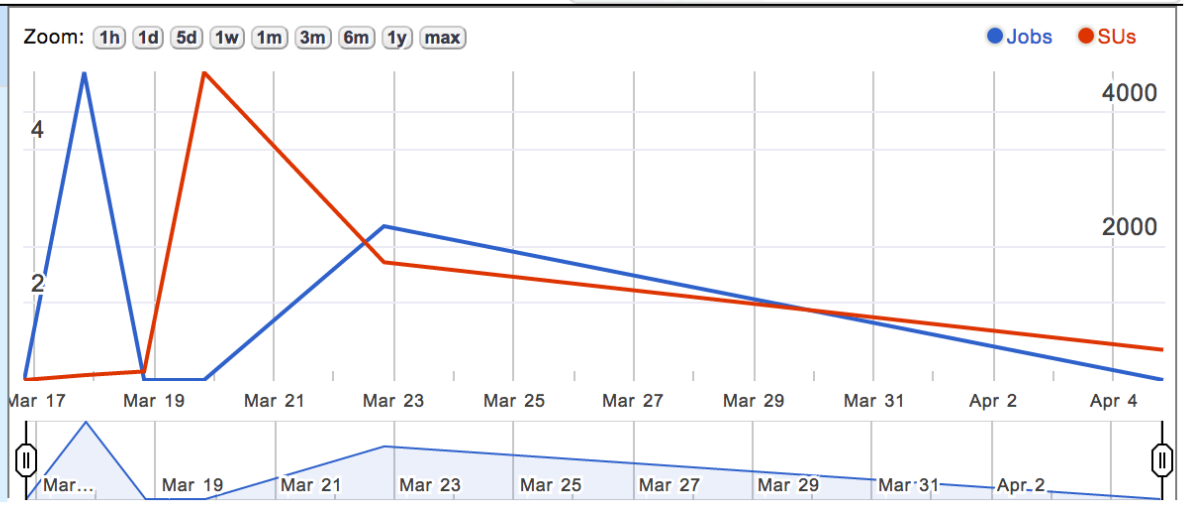


```
slot7@grap-exec0   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:55:41
slot8@grap-exec0   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:52:39
slot9@grap-exec0   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:52:42
slot10@grap-exec1  LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:22
slot11@grap-exec1  LINUX      X86_64 Un
slot12@grap-exec1  LINUX      X86_64 Un
slot1@grap-exec1   LINUX      X86_64 Un
slot2@grap-exec1   LINUX      X86_64 Un
slot3@grap-exec1   LINUX      X86_64 Un
slot4@grap-exec1   LINUX      X86_64 Un
slot5@grap-exec1   LINUX      X86_64 Un
slot6@grap-exec1   LINUX      X86_64 Un
slot7@grap-exec1   LINUX      X86_64 Un
slot8@grap-exec1   LINUX      X86_64 Un
slot9@grap-exec1   LINUX      X86_64 Un
slot10@grap-exec2  LINUX      X86_64 Un
slot11@grap-exec2  LINUX      X86_64 Un
slot12@grap-exec2  LINUX      X86_64 Un
slot1@grap-exec2   LINUX      X86_64 Un
slot2@grap-exec2   LINUX      X86_64 Un
slot3@grap-exec2   LINUX      X86_64 Un
slot4@grap-exec2   LINUX      X86_64 Un
slot5@grap-exec2   LINUX      X86_64 Un
slot6@grap-exec2   LINUX      X86_64 Un
slot7@grap-exec2   LINUX      X86_64 Un
slot8@grap-exec2   LINUX      X86_64 Un
slot9@grap-exec2   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+09:12:12
slot10@grap-exec3  LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:53
slot11@grap-exec3  LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:57
slot12@grap-exec3  LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:57
slot1@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:57:42
slot2@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:58:05
slot3@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:55
slot4@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:54
slot5@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:55
slot6@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:58
slot7@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:55:02
slot8@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:55:03
slot9@grap-exec3   LINUX      X86_64 Unclaimed Idle      0.000 3920  2+08:54:53
                     Total Owner Claimed Unclaimed Matched Preempting Backfill

        X86_64/LINUX    48     0       0        48       0          0        0

               Total    48     0       0        48       0          0        0
grapleadmin@graple-Submit:~$
```

Filter by user: -- All users --

Comet

Oasis



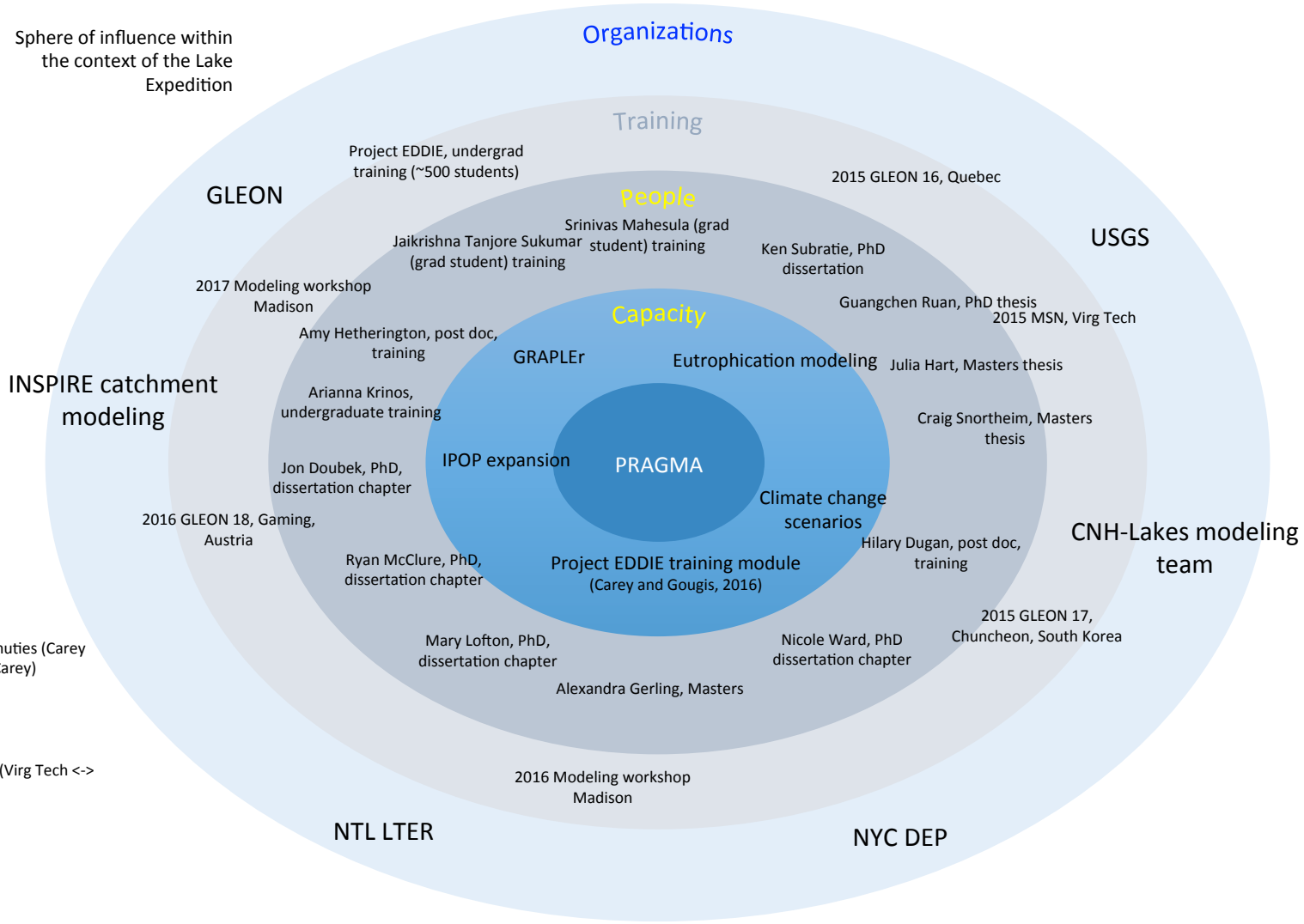
Updates on IPOP and GRAPLEr resources

- 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)
- GRAPLEr lake comparison of nitrogen and phosphorus cycling in GLEON lakes under future climate scenarios (Cayelan Carey, Arianna Krinos, Kate 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!)
- 4 Carey Lab grad students, 1 postdoc, and undergrad extraordinaire are submitting GRAPLEr jobs regularly to advance lake modeling efforts

Sphere of influence within the context of the Lake Expedition



And more:

- Catalyst for proposals, e.g., Smart Commuties (Carey and Figueiredo) and MSB Early Career (Carey)
- NLDAS2 scripting data
- Workflows for data sources
- Sensitivity analysis for GLM-AED
- Optimization/calibration for GLM-AED
- Student exchanges, e.g., Arianna Krinos (Virg Tech <-> UF)

Subratie, K, Aditya S, Carey CC, Hanson PC, Figueiredo R. *In press*. GRAPLEr: A Distributed Collaborative Environment for Lake Ecosystem Modeling that Integrates Overlay Networks, High-throughput Computing, and Web Services. Concurrency and Computation: Practice and Experience.

Ecological Informatics 39 (2017) 10–22

Contents lists available at ScienceDirect

Ecological Informatics

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

Mining lake time series using symbolic representation

Guangchen Ruan[a,*], Paul C. Hanson[b], Hilary A. Dugan[b], Beth Plale[a]

[a] School of Informatics and Computing, Indiana University, 919 E. 10th Street, Bloomington, IN 47408, USA
[b] Center for Limnology, University of Wisconsin-Madison, 680 North Park Street, Madison, WI 53706, USA

Recent publications by the Lake Expedition

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

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

Cayelan C. Carey[1] · Rebekka Darner Gougis[2]

Published online: 22 August 2016

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 large datasets of observations obtained through sensor networks and remote sensing, enabling new analyses and models of ecological phenomena (Hanson 2007; Porter et al. 2005; Weathers et al. 2013). Conducting these analyses and

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