Updates from Resources Working Group

Philip Papadopoulos (UCSD)
Yoshio Tanaka (AIST)

PRAGMA 31

Bangkok, Thailand

Activities at a glance

PRAGMA-RDA in collaboration with Bioscience WG
- Beth/Jason/Gabriel

IPOP/PRAGMA-ENT/Lake expedition
- Kohei/Renato/Paul/Caylan

PRAGMA Cloud
- Shava/Nadya/Aimee/Phil/Steven,et.al.

PRAGMA Website moved to github
- Website: www.pragma-grid.net/
- Source: www.github.com/pragmagrid/pragmagrid.github.io
- Nadya and Phil for accessing request.

Each activity relates with each other by PAGMA Cloud as an anchor.

PRAGMA-RDA-IRRI Demo Phases

Phase 1 (Jan 2016, PRAGMA 30 Manila):

- Discuss how to collaborate with IRRI
  - Problem to address – how to harvest computational experimental results, assign unique identifiers, and share data with other rice researchers

Phase 2 (Mar 2016, RDA P7 Tokyo; Summer 2016, AIST):

- Use 3K rice genomes data subset
  - Small, robust subset of SNPs routinely used for the Tassel pipeline (~4.8 mio SNPs x 3,024 accessions, 13 GB)
- Input datasets and Tassel pipeline bundled into VM
- Defined minimal metadata for data objects
  - Defined data object to contain input files, pipeline dependencies, output files
- Associate two properties with the PID (handle): URL to landing page and pointer to the type definition that describes the minimal metadata

PRAGMA-RDA-IRRI Demo Phases

Phase 3 (Sep 2016, PRAGMA 31):

- Demo in Friday morning: Rice bioinformatics researchers can execute Tassel pipeline using PRAGMA Cloud resources. The new PRAGMA data services are used to identify, download, and faithfully replay the computational experiment. Data objects are also stored in a repository.
- PRAGMA data services: PID Information Types API service (with PRAGMA extension), Data Type Registry service, experimental Handle service, client
- Core functionality achieved, packaging of services for other domains

Phase 3.1 (Sep 2016, RDA P8 Denver):

- Demo same workflow as in PRAGMA 31 as part of adoption report plenary session
- Feedback to RDA PID/DTR working groups

Phase 4 (Apr 2016, PRAGMA 32):

- User testing of service
- Hardening of code and usability improvements in UI and landing pages
- Exploration of contextual provenance inclusion, other PID types, other pipelines

PRAGMA-RDA-IRRI
PRAGMA Cloud: Goals for PRAGMA 32

ntegrate more virtual cluster images and get feedback from early users (Bioscience WG)

- Biolinux VM up and running
- AIST resources for PRAGMA Cloud ready
  - DELL PowerEdge M610, 2 Blades ready, CentOS 6.7
    - Intel Xeon E5620 2.4 GHz 4core * 2 = 16 nodes
    - memory: 24 GB, storage: 600 GB HDD * 2
    - Network: 10GbE: for public/guest networks, 1GbE: for management/storage networks
- Cloudstack drivers for pragmaboot
  - Core functionality implemented
- Tassel pipeline installed
  - Other pipelines are in progress
- Have 3 IRRI users ready to test in upcoming months
  - Feedback for both cluster usage and data service

RAGMA-ENT
pdates since PRAGMA30

- Extension to Thammasat University
  - Wanida and her students have deployed an ENT node at TU connected to ENT backbone using GRE.
- Monitoring & visualization tool [Demo in Thu. afternoon]
  - Boom (a student from Putchong's lab@KU) developed Web-based monitoring & visualization tool for ENT.
- Integration with Rocks and PRAGMA boot [Student Hackathon]
  - The procedure to integrate ENT with Rocks is well figured out. Two student hackathon teams tried to setup a new ENT node within 24 hours by using Rocks and OVS roll.
- Authentication & Secure SDN [Poster]
  - Yamada-san and Kido-sensei (Osaka Univ.) proposed an authentication mechanism for SDN&OpenFlow

IPOP/ENT/Lake Expedition: Expanding resources (by PRAGMA 31)

Lake expedition

- Additional resources from Comet/SDSC and PRAGMA Cloud added to GRAPLEr pool
  - Demo in Friday afternoon.

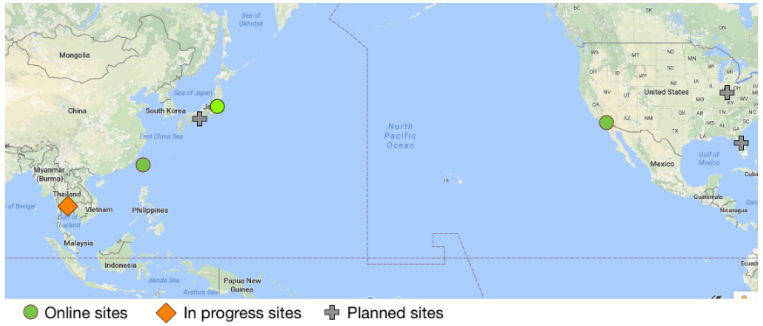
IPOP + ENT

- IPOP virtual networks available for: UF, UCSD, NAIST; still need to deploy for AIST, NCHC

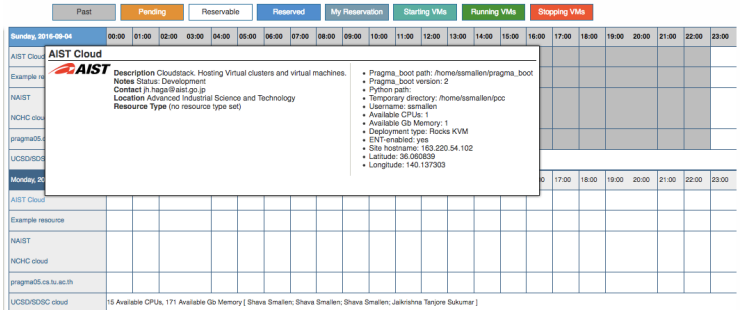
Cloud Scheduler Updates

- Optimized pragma_boot and new driver for Cloudstack
- New Rocks deployment at NCHC
- Updated pragma_boot documentation
- New GRAPLEr image
- New Lifemapper virtual cluster on Comet
- New Clonezilla Live VM and script for converting pragma_boot images
- Two student hackathon topics

PRAGMA Cloud Testbed Sites



PRAGMA Cloud Schduler Screenshot



Cloud Scheduler Related Demos

EMO SESSION #1

Updates for the PRAGMA Cloud Testbed

Making the Universal Image for PRAGMA Boot by Clonezilla

EMO SESSION #3

Dynamic Addition of SDSC Comet and PRAGMA Cloud Nodes to a GRAPL
Pool

Agenda of breakouts

1st day (14:00 - 15:30, 8, Sep.): Discussions for short term goals.

PRAGMA-RDA

PRAGMA-ENT/IPOP/Lake Expedition

PRAGMA Cloud

2nd day (14:30 - 16:00, 9, Sep.): Discussionsfor long term vision.

Presentation: A new data set gathering idea for the Lower Mekong region (Tho Nguyen)

Possible topics for the discussion

- PRAGMA/CENTRA collaboration
- How PRAGMA Cloud hosts/provides data
- High Performance (large-scale) PRAGMA Compute resources
- Future (IoT) Network (e.g. integration of the Internet and M2M)
- Security
- Applicaitons