

Biosciences Working Group

Jason Haga

AIST

PRAGMA 31, Sept 8, 2016

Bioscience Activities

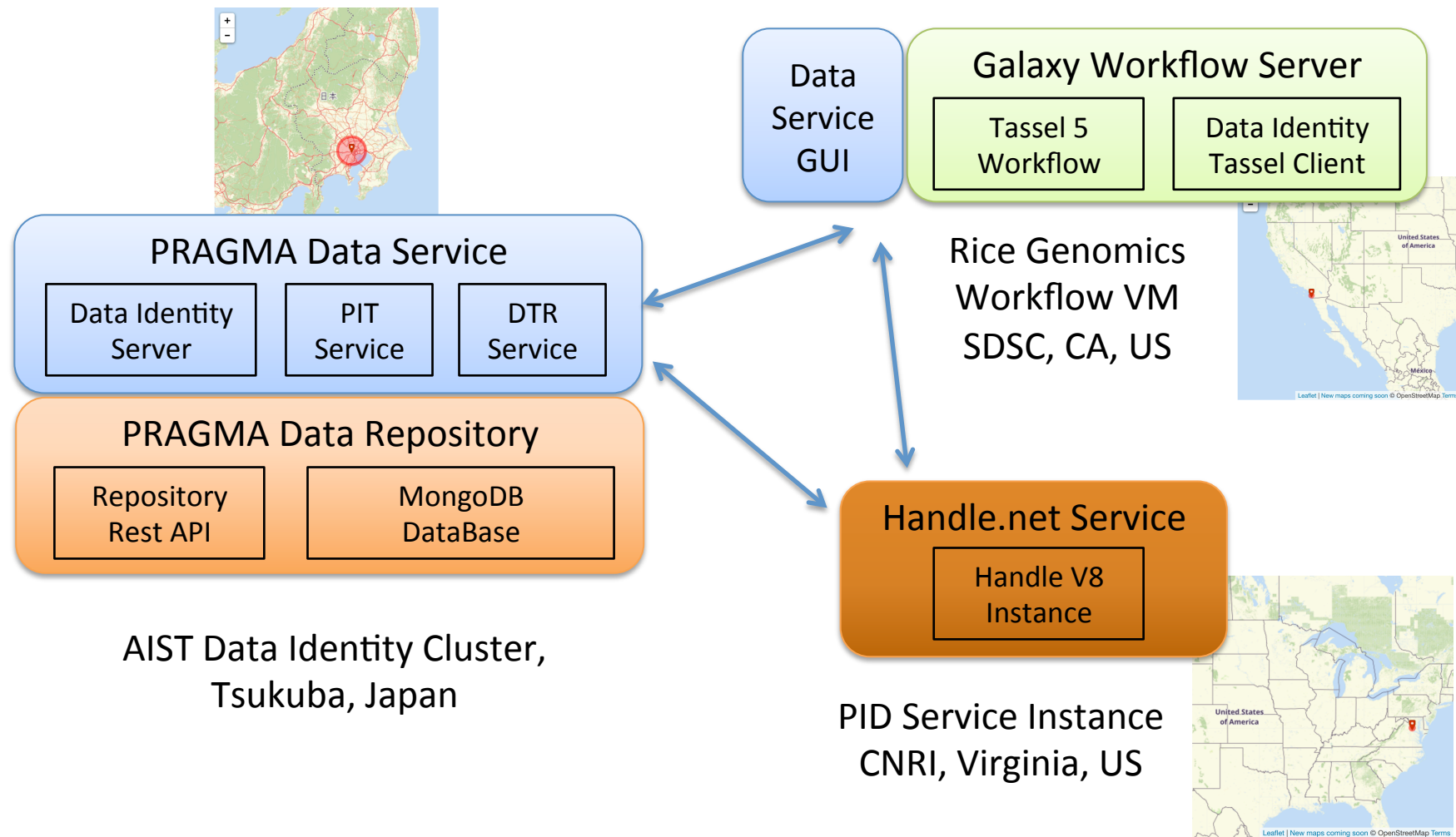
- Genomics Analysis Pipeline
 - Creating resources to enable genomics analysis
- PRAGMA Data Identity Service
 - Rice genomics analysis
- Both driving applications of PRAGMA Cloud
 - Cross-working group interaction
- DOCK in Docker
- Condor application for BLAST

Since PRAGMA30

- Genomics Analysis Pipeline
 - Installation of rice genomics analysis tools such as TASSEL5 (Venice, Ramil)
 - Additional testing of VM (Venice, Gabriel)
 - VM is now accessible via pragmaboot and scheduler (Shava, Nadya)
 - AIST PRAGMA Cloud resources prepared

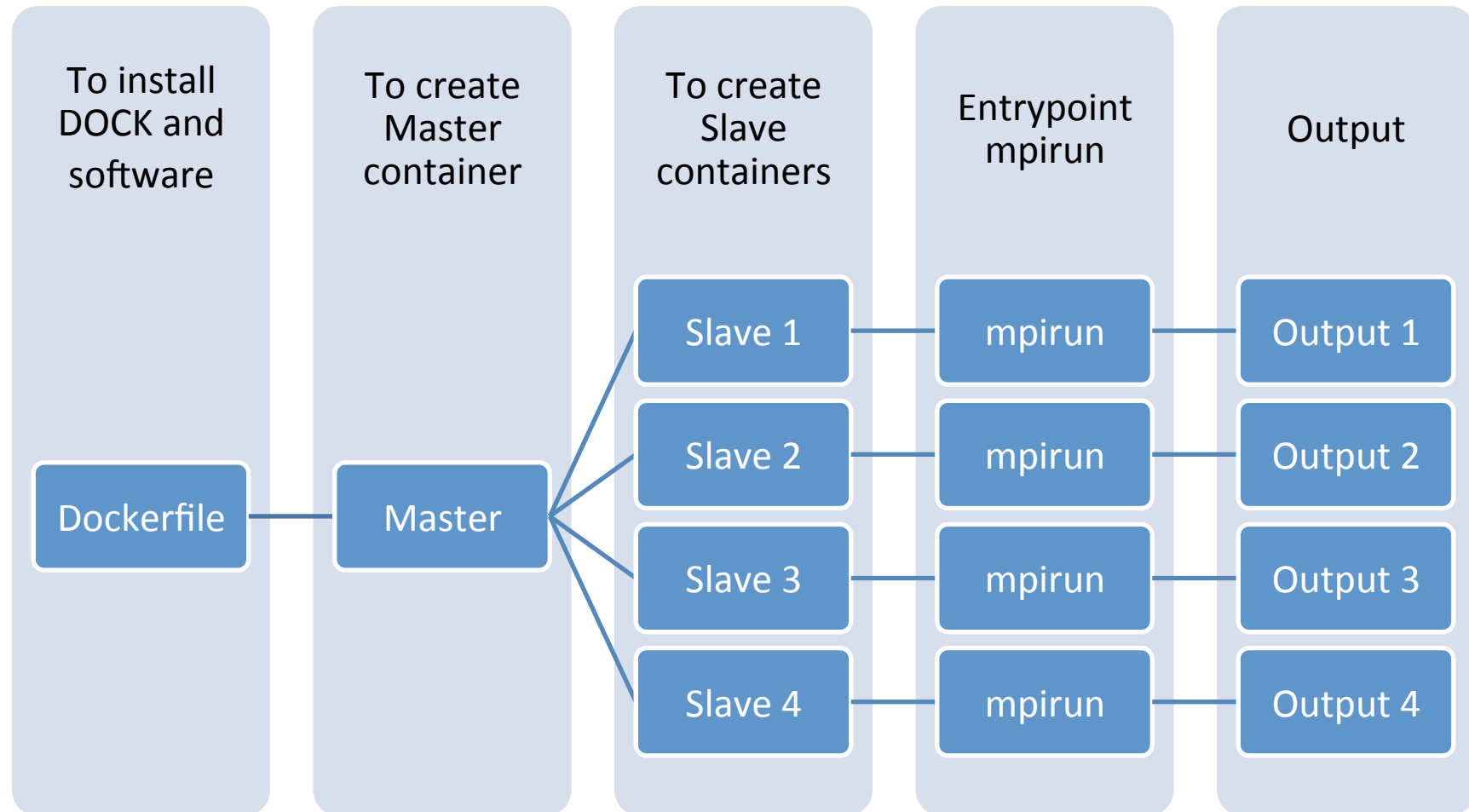
PRAGMA Data Service

(Gabriel, Ramil, Jason, Beth)



DOCK in Docker Containers

(Max, Thomas, Jason, Ichikawa, Weicheng)



DOCK in Docker Containers (cont.)

- **Dockerfile** - To install DOCK and software; generate a dedicated image for DOCK
- **Entrypoint** - File that consists of all actions that will be executed on container
- **Compose** - A tool for defining and running multi-container Docker applications

Performance Evaluation
Total elapsed time (sec)

Number of containers						
	0	2	4	8	16	32
mpirun np 2	3010	3018	----	----	----	----
mpirun np 4	----	----	985	----	----	----
mpirun np 8	----	----	408	416	----	----
mpirun np 16	----	----	211	216	215	----

Condor Computing Application

(Tommy Lam, HKU)

- Utilize **off-hour computers** in library and computer labs in HKU campus
- Demonstration of **Green Computing**
- We are trying to adopt BLAST for condor system, e.g. **BLASTer** (<http://funnelfiasco.com/research/gca13.pdf>)
- BLAST is highly **parallelizable**
 - If you have many query sequences to search against a database, query sequences can be split into different condor nodes for individual BLAST search(easy; BLASTer way)
 - If you have only few query sequences but big database, splitting database into pieces for BLAST is a good option. But this will require MPI or endpoint summary computing of result