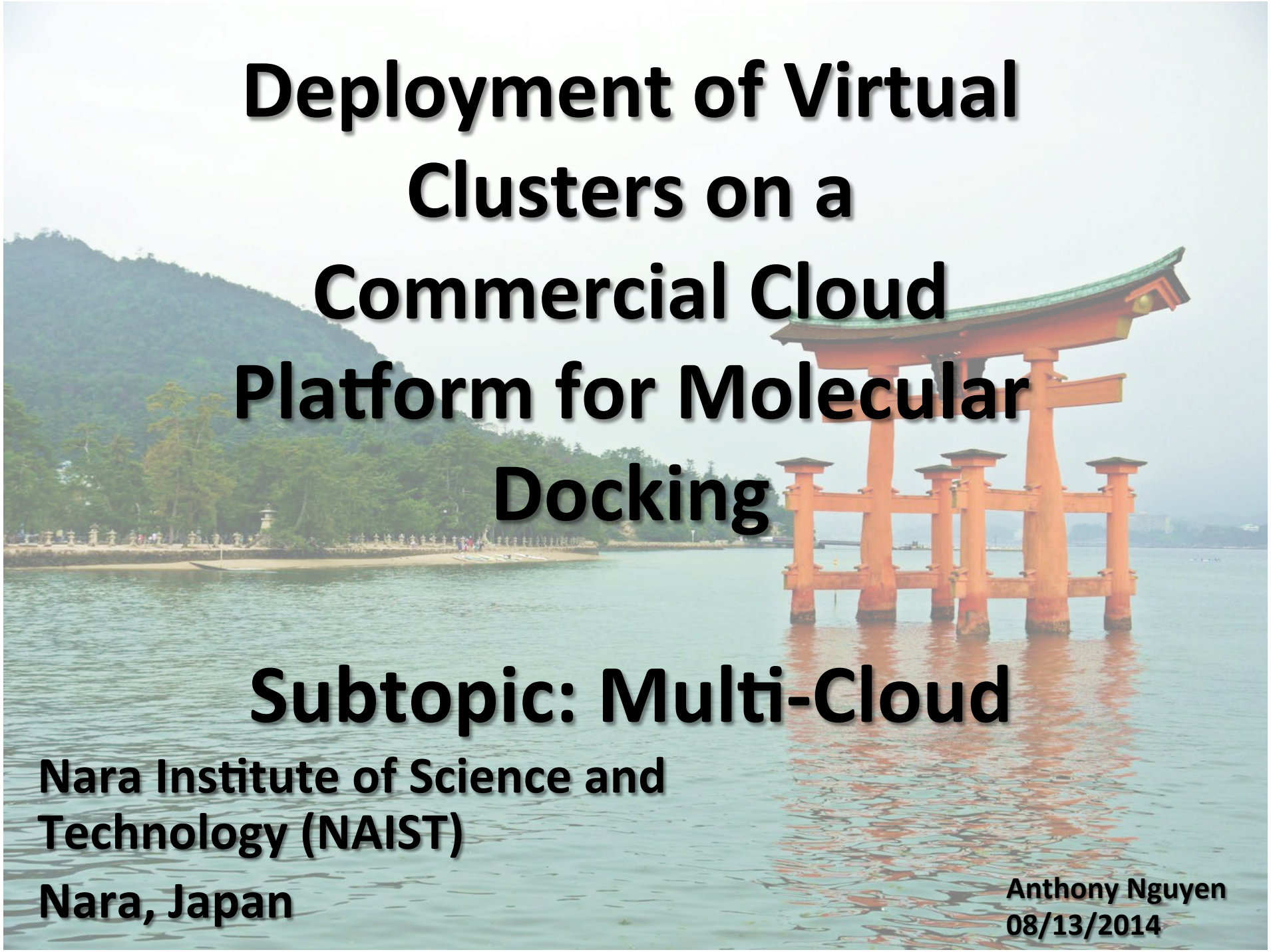

Deployment of Virtual Clusters on a Commercial Cloud Platform for Molecular Docking

Subtopic: Multi-Cloud

Nara Institute of Science and Technology (NAIST)

Nara, Japan

Anthony Nguyen

08/13/2014

Week 8 Progress

- Running Tests
  - With Multi-Cloud Environment fully set up, this week was dedicated to learning and running DOCK tests
  - Ran Tests for Flexible Docking, Rigid Docking, and Amber Score
  - Analysis was done based on observations regarding number of molecules processed and the time to process one molecule by DOCK program on each individual VM

Week 8 Progress

- Test 1: Rigid Test Flexible Parameter off
  - Did a comparison between two tests with different levels of traffic on Clouds.
    - Traffic refers to amount of processes running simultaneously
  - Observed that higher levels of traffic result in slower processing rate per molecule.

Week 8 Progress

- Test 2: Rigid Comparison Test
  - Did a comparison of a Rigid Test on Multi-Cloud Environment with a published result from a Grid Computing Environment
    - Matched as many parameters as possible to standardize comparison.
      - Some variables could not be made identical like DOCK version
  - Result from Multi-Cloud Test in same order of magnitude as that from Grid Computing
    - Tells us Multi-Cloud Environment can perform at least at an equivalent level as Grid Computing while offer much more versatility and flexibility.

Week 8 Progress

- Test 3: Unbalanced Test
  - Change the Cloud Environment increasing the size of FutureGrid by 3 VMs while leaving the others unchanged.
    - 3 on NAIST, 3 on AIST, 6 on FG
  - Distribution of number of molecules processed per VM and the rates to process one molecule per VM were both evenly distributed and similar to balanced test
    - Shows that scaling up one VM does not result in any cloud preference for DOCK jobs.

Final Findings

- Research Results
  - Multi-Cloud environment is able to perform DOCK jobs at least as effectively as a Grid Computing environment (maybe more effectively)
    - While maintaining equivalent performance also has much more flexibility and versatility making it a better environment.
  - Increasing traffic on the cloud environment will naturally increase the rate at which DOCK jobs are performed.
  - Scaling up one cloud to make an unbalanced environment does create any cloud preference
    - Tasks still distributed evenly and processing rates remain roughly the same amongst all the VMs

Future Steps

- Tests
  - Try doing selective high traffic tests, increasing traffic on individual clouds, instead of increasing traffic on all
  - Test on a larger scale
- System Modifications
  - Incorporate Hadoop as a Job Distributor
  - Gain access to and use Commercial Clouds

Cultural Exploration

Acknowledgements

- UCSD Pacific Rim Experiences for Undergraduates (PRIME)
  - Dr. Gabriele Wienhausen
  - Teri Simas
  - Jim Galvin
  - Madhvi Acharya
- Dr. Jason Haga
- Dr. Kohei Ichikawa
- Dr. Mauricio Tsugawa
- PRIME Alumna Haley Hunter-Zinck
- Nara Institute of Science and Technology (NAIST)
- National Institute of Advanced Industrial Science and Technology (AIST)
- URS Ledell
- Japan Student Services Organization (JASSO)
- National Science Foundation
- FutureGrid