GENI Science Shakedown Report
Post GEC22 - April 2015

The GENI Science Shakedown project has made significant progress for the time period including GEC21 and GEC22. The high-level goal of this period was to port the ADCIRC Storm Surge modeling workflow to GENI and scale the computation to the largest slice possible (within reason). The outcome goals of this period was to have a high-profile demonstration of the large ADCIRC slice at GEC21 and/or GEC22.

## GEC21

In the months leading up to GEC21 we were able to accomplish the following tasks:

**Port ADCIRC (ExoGENI)**: We ported the existing ADCIRC software to ExoGENI. This included porting the MPI portion of the application, not the full Pegasus workflow. This task involved installing the appropriate software into a VM, creating a snapshot of the image, and creating a RSpec request that recreates the slice. At that point, ADCIRC on GENI was limited to individual runs that needed to be executed by-hand on small (32 core) groups of ExoGENI nodes.

**Port ADCIRC (InstaGENI)**: Instead of porting ADCIRC directly to InstaGENI, we decided to develop a process by which anyone could import ExoGENI images into InstaGENI. We then were able to import the ExoGENI ADCIRC image into InstaGENI and create RSpec request files that created ADCIRC slices on InstaGENI. These slices have the same functionality as the ExoGENI slices and were limited to individual runs that needed to be executed by-hand. It was not possible to create a generic process for importing images from ExoGENI to InstaGENI. Each image has its own details that must be handled properly. We were only able to document the process of importing ExoGENI's standard Centos6 image into InstaGENI. It should be noted that this was only possible with a lot of help from the Emulab team (Leigh Stoller) and other GENI members (Ezra Kissel).

**Scale ADCIRC on GENI:** One goal of this project was to test GENI's ability to support domain science application at the largest scale possible. Toward this goal we deployed set of ADCIRC slices on GENI that could run 6 ensemble members simultaneously. The slices included one management VM and 6 sets of compute VMs on 6 different GENI racks (3 EG, 3 IG). The sets of compute VMs each contained a total of 32 cores (8 VMs with 4 cores/VM).

**Shakedown GENI Stitching:** As part of the scaling exercise we ran into many issues with GENI stitching between sites. An additional contribution of this project is that we pushed GENI stitching past its limits and reported our experiences to the appropriate developers. As a result, GENI stitching is much more stable.

**GEC21 Demo:** The most noticeable outcome of our work up to GEC21 was the GEC21 plenary demo. In the demo we showed the live creation of a large ADCIRC slice and started the ADCIRC execution on that slice. The slice included 38 VMs (152 compute cores) across 6 GENI racks (3 EG, 3 IG) and ION.

## GEC22

In the months leading up to GEC22 demo we were able to accomplish the following tasks:

**ADCIRC Workflow**: We developed at Pegasus workflow to manage the 22 member ADCIRC ensemble. This workflow was incorporated into the slice descriptions and was executed on GENI slices.

**Intelligent Data Movement Service (IDMS)**: The ADCIRC workflow includes the transfer of approximately half a terabyte of data that needs to be transferred to/from the compute hosts. As part of the GEC22 demo we incorporated GENI's IDMS service to handle the data transfer where possible.

**Further Scale of ADCIRC on GENI:** For the GEC22 demo we further scaled the application to include additional GENI resources.. This version of the ADCIRC slices could run 10 ensemble members simultaneously. The slices included one management VM and 10 sets of compute VMs on 10 different GENI racks (5 EG, 5 IG).

**GEC22 Demo:** The most noticeable outcome of our work upto GEC22 is the GEC22 plenary demo. In the demo we used a previously created large ADCIRC slice to run the full ADCIRC workflow. The slice included 11 GENI sites (1 ensemble manager, 5 EG compute sites, 5 IG compute sites) This topology included 92 VMs (368 cores), 10 inter-domain VLANs, 1 TB iSCSI storage. The HPC compute nodes allone accounted for 80 compute nodes (320 cores) from 10 sites. In addition, each VM in this demo included OML monitoring that fed into a custom visualization created by the GPO.