UCSD/Qualcomm Institute GPU Challenge - Nautilus and CHASE-CI services: At-scale machine learning and analysis with NASA MERRA data

Scott L. Sellars[1], John Graham[1], Dima Mishin[1], Tom DeFanti[1], Larry Smarr[1], Joulien Tatar[2], Eric Shearer[3], and Phu Nguyen[3]

[1]California Institute for Telecommunications and Information Technology (Calit2), UCSD, La Jolla, CA
[2]Office of Information Technology, University of California, Irvine, Irvine, CA
[3]Center for Hydrometeorology and Remote Sensing, University of California, Irvine, Irvine, CA

Rapid object segmentation for evolving time and space earth science data is challenging. Often in the earth sciences, objects of interest (e.g., rain clouds, flash floods, droughts) are not clearly defined and change rapidly and dynamically in time. Specialized algorithms are needed to identify, locate, and track these types of phenomena. In this GPU Challenge demonstration, we use a Jupyter Notebook connected to Nautilus and CHASE-CI services to use the new Kubernetes GPU cluster and TPU resources to do rapid object segmentation by applying a machine learning approach: Flood-Filling Networks (FFN) [*Januszewski et al., 2016*] to the NASA MERRA v2 data stored on a UCSD FIONA, accessible using a THREDDS server on the Pacific Research Platform (http://its-dtn-02.prism.optiputer.net:8080/thredds/catalog.html). These resources provide the capability for an automated algorithm development and deployment platform across the CHASE-CI kubernetes GPU cluster.

The goal is to take full advantage of the 100+ GPUs, high-speed network provided by the Pacific Research Platform, and a Redis database for rapid data segmentation and visualization. In addition, we compare the performance of this FFN results to the CONNected objeECT, or CONNECT algorithm [*Sellars et al., 2013, 2015, 2017*] that it keeps track of the entire life-cycle of a detected object, from the genesis to the termination location by "connecting" pixels in time and space, and tracking the system to be analyzed as a "4D" object. During the demonstration a description of all system developments and innovations are provided, including the use of CHASE-CI services: Docker images stored in the Nautilus Docker image repository (https://gitlab.nautilus.optiputer.net/connect), Nautilus RocketChat, and Nautilus Kubernetes Cluster.