Draft Report to NSF on Pacific Research Platform Application Usage

Prepared by Larry Smarr, Distinguished Professor Emeritus, UC San Diego

December 10, 2020

The PRP team is able to track the computing (CPU and GPU) usage of each of the 400+ namespaces in PRP's Nautilus Kubernetes hypercluster, which connects nearly 30 campuses and national laboratories. The PRP Nautilus distributed cyberinfrastructure hosts over 7000 CPU-cores and over 500 32-bit GPUs.

Below is an outline of some of the PRP Nautilus namespace application users over the first nine months of 2020 [hereafter referred to as "in 2020"].

- **Data analysis for large scale scientific instruments** (LHC, IceCube, LIGO, ...)
  - In the original PRP proposal, data analysis of both the CMS and Atlas Large Hadron Collider (LHC) were anticipated and have been a major user of Nautilus.
  - The NSF-funded LIGO gravitational wave observatory was mentioned in the original proposal, but after the first black hole collisions were observed five years ago, the use of Nautilus for data analysis by LIGO researchers has steadily increased and in 2020 LIGO has begun to make significant use of Nautilus GPUs. Also, the PRP cache at the University of Amsterdam significantly reduces LIGO transatlantic network traffic.
  - Not included in the original proposal, the NSF South Pole IceCube neutrino observatory has become the major namespace application users of the Nautilus GPUs, consuming roughly half of all Nautilus GPUs throughout 2020. In Q4 of PRP Year 5, Nautilus was the third largest provider of GPUs to IceCube, in spite of its unanticipated late arrival as a PRP application user. This enables the computation of parallel photon paths in ice to help increase the pointing accuracy of the IceCube observatory.
- **Community scientific database generation and distribution, particularly in biology** (OpenForceField, Folding@home, biological modeling)
  - Two other major science projects, that were not in the original PRP proposal, became major users of PRP CPUs and GPUs during the COVID-19 pandemic, Folding@home and OpenForceField. The global Folding@home effort is carrying out highly distributed computation of molecular dynamics to look for new druggable sites on the spherical spiked RNA Virus SARS-CoV-2, which causes COVID-19. During 2020, Folding used over one million Nautilus CPU-core hours CPUs, adding ours into the rapidly expanding collection of computers that eventually made Folding more powerful than all the Top500 supercomputers combined.
  - The Open Force Field Initiative (openforcefield.org) "discovered" PRP in 2020 and used ~1.8 million CPU-hrs to generate quantum chemical training data for their next generation of open-source molecular models for druglike molecules, which are in turn used to train simulation models for applications in molecular design, biophysics, biochemistry, and drug discovery, including the COVID-19 efforts on Folding@home. The PRP leadership team received an unsolicited letter from the OpenForceField team on April 20, 2020 stating that "the PRP is superbly well-architected for our needs."

- Besides these community biology projects, one of the larger individual faculty Nautilus CPU/GPU application users is the Professor Mark Alber group at UC Riverside, whose team computes models of biological processes, such as: early development, epithelial rounding, chemical signaling, fibrin fiber dynamics, and blood clot formation. The *markalbergroup* namespace used about ¼ million CPU core-hours and ¼ million GPU-hours in 2020, one of the examples of balanced requirements for both CPUs and GPUs to carry out their scientific applications.

- **Dramatic speedup of networked workflows for hazard management** (atmospheric rivers, wildfire prediction)
    - An early example of the use of the high speed PRP networks to speed up a large data science analysis was the connection via PRP of two atmospheric water research groups (UC Irvine's Center for Hydrometeorology and Remote Sensing and UCSD's Center for Western Weather and Water Extremes) by researcher Scott Sellars. This workflow ingested NASA satellite global precipitation data, then carried out machine learning on precipitation system evolution data objects and used them to study the often devasting atmospheric rivers that can bring torrential rain to the Southwest United States. By careful optimization of each step of this workflow on the PRP components (network, CPUs, GPUs), Sellars was able to demonstrate a 500x speedup (from 19.2 Days to 52 minutes!). This converts the computational pipeline from a historical study of a major precipitation event to a prediction capability, essential for mitigating these hazardous meteorological events.
    - Wildfire observation and prediction system. Over the last few years, the PRP has connected the High Performance Wireless Research and Education Network (HPWREN) data servers with the San Diego Supercomputer Center Comet supercomputer. The real-time meteorological data from weather sensors on the HPWREN radio towers is fed into the wildfire software system developed under the NSF-funded WIFIRE grant to enable faster-than-real time predictions of the path of a wildfire. Thus, four NSF-funded grants are sewn together (HPWREN, WIFIRE, PRP, and SDSC's Comet) to create a nationally unique wildfire simulation system whose output is fed directly into the Calfire and local first responder incident command centers. In addition, by interconnecting the HPWREN camera image servers at UCSD, SDSU, and UCI over the CENIC optical network, the public-facing web servers can now continue to provide instant response as the public demand for those camera images increases rapidly by 10-100x when a wildfire occurs.

- **Machine learning research** (image and language analysis, self-driving cars, deep learning)
    - Because the PRP companion Cognitive Hardware and Software Ecosystem Community Infrastructure (CHASE-CI) NSF grant supplied a number of the PRP Nautilus GPUs, we have built a community of machine learning (ML) and artificial intelligence (AI) users. Together, this community's use of PRP's Nautilus GPUs is in aggregate equal to the IceCube data analysis mentioned above. This broad diversity of applications includes: refinement and acceleration of ML methods based on point clouds for improving self-driving cars, using AI approaches to uncover how genetic changes enhanced human brain architecture and computing capacity during primate evolution, integrating genomics and deep learning for new insights into cancer diagnostics, developing slim

and accurate deep neural networks which are crucial for real-world applications of embedded systems in drones and smart phones, using deep learning to do image segmentation on 3D Teravoxel-scale electron microscopy volumes, and unsupervised domain adaptation for semantic segmentation and multi-domain learning for several classification datasets.