MULTIMESSENGER ASTROPHYSICS WITH THE BLUE WATERS SUPERCOMPUTER

Allocation: Innovation and Exploration/1,000 Knh
PG: Eliu Huerta1, Fajardo3, Stuart Anderson4, Peter Couvares4, Joshua Willis4
1University of Illinois at Urbana-Champaign
2National Center for Supercomputing Applications
3University of California, San Diego
4California Institute of Technology

EXECUTIVE SUMMARY

We developed a novel computational framework that connects Blue Waters, the NSF-supported leadership-class supercomputer operated by NCSA, to the Laser Interferometer Gravitational-wave Observatory (LIGO) Data Grid via Open Science Grid technology. This work represents the first time Open Science Grid, containers, and Blue Waters were unified to tackle a scientific problem at scale. This new framework has been used during LIGO’s second discovery campaign to run the most computationally demanding LIGO workflows on Blue Waters to accelerate discovery in the emergent field of gravitational wave astrophysics, and to validate the first gravitational wave detection of two colliding neutron stars with the LIGO and Virgo gravitational wave detectors. This discovery that marks the beginning of multimessenger astrophysics (MMA). Supporting LIGO data analysis workflows concurrently and Blue Waters, the NSF-supported leadership-class supercomputer was unified to tackle a scientific problem at scale. This framework can be readily used to run other scientific workflows on the Blue Waters supercomputer if they meet the following requirements: they are a good match to the OSG infrastructure, the software can be containerized, and a workflow manager can be used to monitor the workflow from end to end, i.e., Pegasus, Swift, etc. This is a minimal set of requirements that may be easily met by existing OSG users, who may already use portable, self-contained software that could be containerized.

METHODS & CODES

The Open Science Grid (OSG) provides federated compute resources for data-intensive research in a variety of science areas [2]. OSG targets typical high-throughput workloads consisting of spatially small, loosely coupled science jobs that are executed on any of the participating resources providing clusters [3]. We used this flexibility to target high-throughput computing workloads on Blue Waters. LIGO currently uses the Pegasus Workflow Management System [4] as a layer on top of DAGMan (Directed Acyclic Graph Manager) to manage dependencies. DAGMan is provided by HTCondor to enforce dependencies among jobs in large workflows, and reliably restart workflows from point of failure. PyCBC [5] is one of LIGO’s most computationally intensive gravitational wave search pipelines and the only production systemwide, parallel file systems are visible inside the container; that MPI can be used; and that Blue Waters’ security policy is enforced on the container.

To validate our results, we first ran a small PyCBC workflow on OSG facilities using the data set utilized by GitHub Travis CI tests on LIGO clusters. This data set and the results obtained from this analysis have been thoroughly cross-checked using LIGO and OSG resources. Having a baseline for comparison, we ran a PyCBC workflow on Blue Waters using the same validation data set and thoroughly checked that the results reported in both independent analyses were identical. Theretofore, we repeated the same exercise running 10 times larger PyCBC workflows both on OSG and Blue Waters, and confirmed that the results were consistent. Upon confirming that our computational infrastructure works in a stable manner and that we were able to accurately reproduce results obtained with OSG resources, we stress-tested this new framework with several production-scale workflows. The computational framework used for these studies is presented in Fig. 1.

RESULTS & IMPACT

This framework marks the first time convergence was reached on the Blue Waters supercomputer and exhibited the flexibility and interoperability of NSF cyberinfrastructure to enable and accelerate scientific discovery. We have used this novel computational framework at scale to validate the gravitational wave detection of two colliding neutron stars with the LIGO and Virgo detectors [7], as shown in Fig. 2.

Figure 1: Use of Shifter to run LIGO workflows on Blue Waters.

Figure 2: Open Science Grid compute resources used for large-scale gravitational wave data analysis. The chart shows the first time Blue Waters was used at scale as an Open Science Grid compute element, which corresponds to the gravitational wave discovery of two colliding neutron stars by the LIGO and Virgo Detectors.