MULTIMESSENGER ASTROPHYSICS WITH THE BLUE WATERS SUPERCOMPUTER

Allocation: Innovation and Exploration/1,000 Knh
PI: Eliu Huerta[1]
Collaborators: Roland Haas[2], Daniel S. Katz[2], Timothy Bouvet[2], Jeremy Enos[2], William T.C. Kramer[2], Hon Wai Leong[2], David Wheeler[2], Edgar Fajardo[3], Stuart Anderson[4], Peter Couvares[4], Joshua Willis[4]

[1]University of Illinois at Urbana-Champaign
[2]National Center for Supercomputing Applications
[3]University of California, San Diego
[4]California 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 with highly parallel numerical relativity simulations is the most recent success and most complex example of successfully achieving convergence with the Blue Waters supercomputer.

RESEARCH CHALLENGE

Future gravitational wave discovery campaigns will be longer and will involve more detectors, thereby requiring additional compute resources. Furthermore, to support urgent needs such as the detection of gravitational waves that are accompanied by emission of light and neutrinos, LIGO will need to rely on resources beyond those available for its normal processing to validate these discoveries. Other challenges have to do with NSF and HPC infrastructure. Computationally demanding workflows allow Blue Waters to increase cluster utilization and throughput. The computational tasks required for this work is good utilization of backfill for unused Blue Waters nodes. MMA require the interoperability of NSF cyberinfrastructure resources and show how large projects can benefit from making use of existing resources rather than having to build their own custom solutions for all possible needs.

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 pipeline that currently runs on OSG. It has been thoroughly tested by LIGO on a number of different clusters. However, all of them used a variant of the RedHat or Debian operating systems on their compute nodes, where specific versions of these operating systems were used to certify the software stack. This presented a challenge, as Blue Waters does not operate on these tested Linux variants, but rather on a lightweight Linux variant based on SUSE. To overcome this challenge, we adopted Shifter [6] as Blue Waters' container solution. Shifter accepts Docker image files and converts them into a disk image suitable for concurrent use by multiple Blue Waters compute nodes. Further, Shifter ensures that the 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.

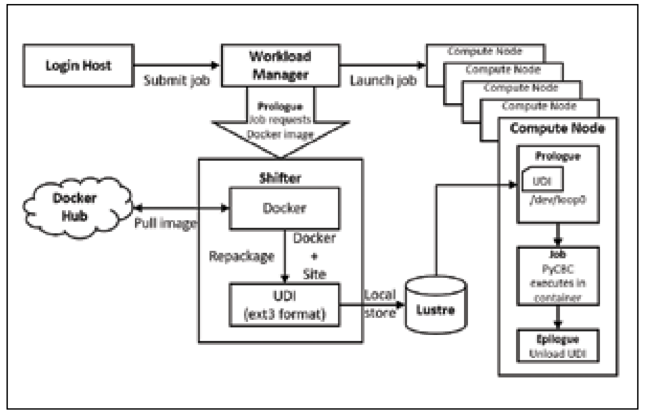

Figure 1: Use of Shifter to run LIGO workflows on Blue Waters.

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 LDG clusters. This data set and the results obtained from this analysis have been thoroughly cross-checked using LDG 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. Thereafter, 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 work 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.

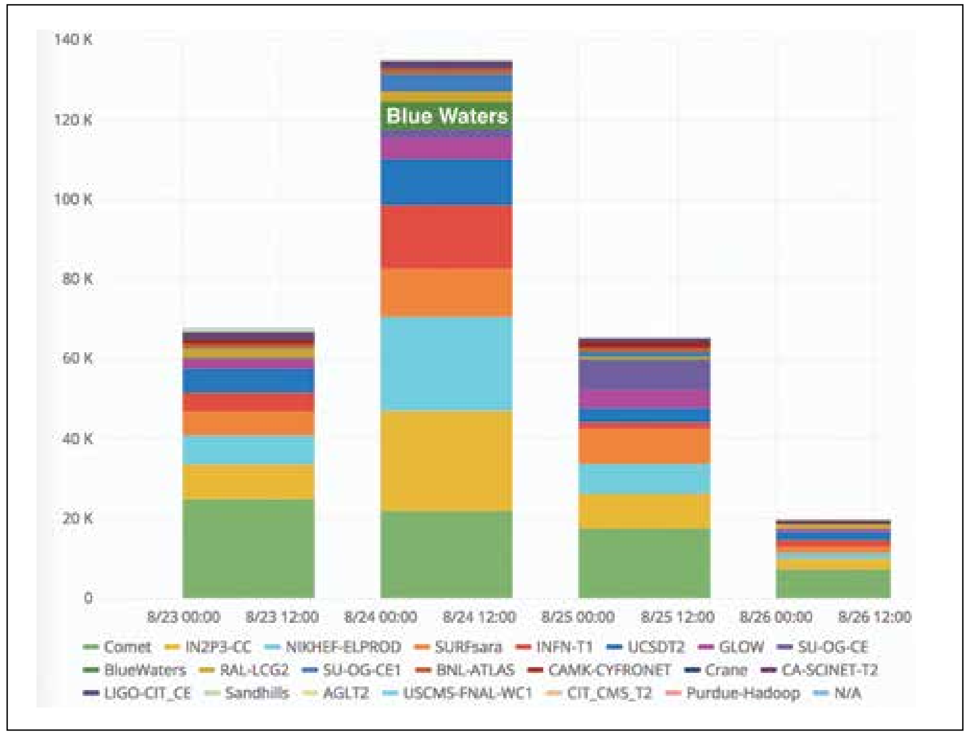

Figure 2: Open Science Grid compute resources used for large-scale gravitational wave data analysis. The chart shows the first time Blue Waters was used at scale as an Open Science Grid compute element, which corresponds to the gravitational wave discovery of two colliding neutron stars by the LIGO and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

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.

WHY BLUE WATERS

The Blue Waters supercomputer is ideally suited to facilitate large-scale gravitational wave data analysis because the large number of independent jobs in these analyses can quickly be run using the reasonably large set of otherwise unoccupied nodes through backfill. Blue Waters staff assisted us throughout the development and exploitation of this framework at scale for gravitational wave discovery.

PUBLICATIONS & DATA SETS

Huerta, E.A., et al., BOSS-LDG: A Novel Computational Framework that Brings Together Blue Waters, Open Science Grid, Shifter and the LIGO Data Grid to Accelerate Gravitational Wave Discovery. *eScience*, (2017), DOI:10.1109/eScience.2017.47.