2021

R&D 100 WINNER

Optimizing Workflow with Flux

TODAY'S supercomputers enable researchers to simulate phenomena and investigate data across vast spatial scales—from the smallest atoms to the largest objects in the universe. Scientific applications in the exascale era will grow more diverse and complex as researchers seek to accomplish simultaneous tasks, such as combining large-scale simulations with in situ visualization, data analysis, machine learning, and artificial intelligence. Such efforts create advanced workflows that span not only individual supercomputers but multiple high-performance computing (HPC) clusters to dramatically increase the scale of computation and data analysis.

Resource management and workflow scheduling software help monitor hardware availability and assign access to HPC systems as researchers submit requested jobs, which encapsulate dependent and independent tasks from a workflow. However, traditional resource managers and workflow managers cannot keep pace with increasing system scales

and interplays such as those occurring between multiple compute clusters and file systems nor are they designed to address converged computing, in which HPC and supercomputing sites leverage cloud infrastructure to improve performance, portability, and accessibility of scientific applications. A framework is required to manage resources and complex workflows efficiently and seamlessly while adapting to various operating scales and infrastructure ranging from exascale supercomputers to medium-scale HPC clusters to on-demand cloud platforms.

Honored with a 2021 R&D 100 Award, Flux is a scalable, flexible next-generation workload management framework that meets this need—maximizing resource utilization while allowing scientific applications and workflows to run faster and more efficiently. Developed in collaboration with university partners, Flux also enables new resource types, schedulers, and services to be deployed at data centers as they continue to evolve. Former Livermore computer scientist and principal investigator for Flux, Dong H. Ahn,

says, "Flux is a low-level piece of software that makes high-level computing possible."

Beyond State-of-the-Art

Flux's first major advantage over competing technologies is the software's hierarchical approach to resource management and workflow scheduling. Flux breaks down bundled user requests into subtasks and then manages each subtask with an individualized sub-scheduler. For example, a workflow requiring high throughput of many small jobs could submit a larger, longer running parent job and use a specialized scheduler to maximize small sub-job throughput. Higher throughput allows users to run larger job ensembles, generating more simulation data. This hierarchical design allows any workflow to use Flux as a component layer of a larger workflow.

Next, Flux's graph-based scheduling uses directed graphs to model system resources, check states, and manage allocations to manage resources with complex relationships. Users can spin up their own personal Flux instance and fine-tune the Flux scheduler to fit their needs rather than be tied to specific schedules and settings. Becky Springmeyer, Livermore Computing (LC) division leader and Flux project leader, says, "Users benefit from pluggable schedulers with deeper knowledge of network, I/O (input/output), and power interconnections, and the ability to dynamically shape running work."

Finally, Flux offers portability to different computing environments, particularly
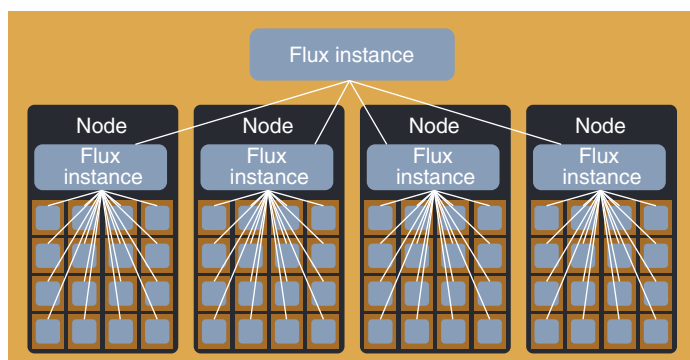
Any Flux instance can spawn child instances to aid in scheduling, launching, and managing complex compute job sequences until the lowest level instance manages a single core.

Development team for Flux: (from top left) Thomas Scogland, Albert Chu, Tapasya Patki, Stephen Herbein, Mark Grondona, Becky Springmeyer, Christopher Moussa, Jim Garlick, Daniel Milroy, Clay England, Michela Taufer, Ryan Day, Dong H. Ahn, Barry Rountree, Zeke Morton, Jae-Seung Yeom, and James Corbett.

new HPC systems with heterogenous architectures, such as Livermore's upcoming exascale system El Capitan—projected to be the world's most powerful supercomputer when fully deployed in 2023. Workflows can also be enabled on other computing systems, at remote locations, on laptops, and in the cloud. Workflows requiring multiple sites no longer need to code to each site-specific scheduler. Instead, they can rely on Flux to handle the nuances of each site.

Flux was the brainchild of LC systems software experts, who recognized that existing resource managers and workflow managers would not handle the scale and complexity of future HPC systems. The team initiated Flux's foundational project in 2014 and expanded the research to include other researchers from the Laboratory as well as the University of Delaware and the University of Tennessee at Knoxville. Active involvement among researchers, operational specialists, users, and vendors in the design phase ensured a workable framework for both researchers and developers. "We tried to create an environment in which the production side and the research side could talk to each other. A tremendous level of trust exists between the two teams, and they truly understand that they need one another to produce the best possible product," says Springmeyer.

Real-World Applications

Flux has been instrumental to several scientific discovery projects, including efforts geared toward combatting the COVID-19 pandemic. In 2020, to improve understanding of the disease and help develop response strategies, the Department of Energy formed the National Virtual Biotechnology Laboratory (NVBL), a consortium of several national laboratories. NVBL urgently needed computing cycles to model the spread of COVID-19 but did not have time to tailor coding to different schedulers at each supercomputing site. Instead, NVBL researchers used Flux to program their complex workflows and manage the intricacies of running jobs at each site while simplifying site-specific resource managers.

For Livermore's COVID-19 antiviral small molecule project, a multidisciplinary team was tasked with developing a highly scalable, end-to-end drug design workflow to produce potential COVID-19 drug molecules for clinical testing. When creating an end-to-end solution based on the existing components presented workflow limitations, Flux's fully hierarchical framework allowed researchers to overcome scalability issues to continue drug design research. Flux has also been a foundational component of Livermore's Merlin workflow, enabling

machine learning–ready HPC ensembles; the American Heart Association Molecular Screening workflow; and the Laboratory's Autonomous Multi-Scale Strategic Initiative advancing embedded machine learning for smart simulations.

Flux is open-source software available on GitHub that can be used freely by HPC centers, cloud providers, and users around the world. The Flux team continues to harden the framework ahead of widespread production deployment at the Laboratory while adding enhancements to support cloud computing environments such as Kubernetes. Research collaborations with other academic institutions may yield greater improvements, and industry collaborations to develop best practices include Amazon Web Services, IBM, and Red Hat OpenShift. As Ahn says, winning the R&D 100 Award "is just the starting point."

—Stephanie Turza

Key Words: Flux, high-performance computing (HPC), National Virtual Biotechnology Laboratory (NVBL), open-source software, R&D 100 Award, workflow manager.

For further information contact Dan Milroy (925) 424-4419 (milroy1@llnl.gov).