Embracing the New Multi-level Compute Era: A Common Workflow Toolkit that Just Works

Dong H. Ahn, Daniel Laney, Kyle Chard, Shantenu Jha, Todd Munson,

The latest trend in high performance computing (HPC) is characterized by a convergence of traditional high-end and cloud computing. While there is a flurry of buzzwords about this particular convergence, our position is to view this as the first step towards an ultimate destination: i.e., a seamless compute infrastructure being built around the concept of multi-level compute hierarchy. At the top of this hierarchy will still lie the largest high-end HPC systems like the DOE's supercomputers; at the next level, cloud computing farms; and the bottom will be characterized by so-called edge computing. In the next two to three decades, we are expected to witness increasing numbers of scientific workflows that will not only be run on one compute level, but simultaneously and seamlessly run across all of these different levels.

Towards this goal, a revolution will be accelerated across each and every level. HPC will be radically democratized and become ubiquitous in both research community and the private sector, analogous to the uptake of AI and analytics technologies. While we do not know the exact set of technologies that will be in vogue, the HPC and AI ecosystems will also increasingly interplay in a variety of ways. Other trends will include, but not be limited to:

- Continuing growth of on-premises large computing centers on the cutting edge of the HPC capabilities at DOE labs and private sector centers;
- Extreme heterogeneity in hardware (Quantum, accelerators, etc.);
- Ubiquitous cloud resources, from basic analytics compute, to heterogenous high performance hardware presumably adopted partly from DOE HPC initiatives;
- Vast numbers of performant, high distributed localized HPC compute nodes in the form of desktops, embedded systems, even cell phones;
- Billions of devices and sensors in all manner of domains begin leveraged for everything from national security to making better products.

The core feature of this future is the complexity and size of this ecosystem, and potential of almost anyone to leverage HPC to further their goals, whether business, research, or national security.

How can the DOE facilitate this playing field where even small companies, or small colleges, or enterprising individuals can leverage HPC technologies? Perhaps more importantly, how can the DOE benefit from this so-called multi-level compute revolution? The first question is centered around the concept of *spin-off* whereby the DOE contributes to the success at the lower levels of compute. The second question is about *spin-back* whereby the DOE harvests its investments back for its own mission-critical sciences. We believe the DOE can realize the full potentials of this revolution only when its infrastructure software can *uniformly* and *ubiquitously* underpin the computation performed across all levels of compute in the research, government and private sectors alike.

Among the infrastructure software, adapting our resource and job management software (RJMS) and workflow management software (WMS) systems for this era will be particularly critical. Here, the key challenges will be enabling millions of potential users of this vast technology landscape to leverage the resources it provides. For example:

- Building applications and workflows that span local compute to distance cloud;
- Scheduling a hardware landscape of such complexity; and
- Connecting vast quantities of data with computation across all levels.

More specifically, the RJMS and WMS systems must provide common functionality, portability, scalability and resiliency from the future exascale desktop machine, to a vast distributed set of cloud resources, to high-end supercomputing resources. Because each computing environments and use cases will be vastly different, a *one-size-fits-all* approach will simply not work. We rather need a toolbox consisting of key management components such as workflow orchestration, specification, runtime, system and user-level resource management and batch-job scheduling, which can function as a stand-alone tool but also can be seamlessly composed together through well-defined interfaces.

The key challenge will be to meld the typically underfunded, yet cutting edge, efforts in DOE and other government labs, with the large resources and teams that private sector cloud vendors currently bring to bear for the workflows of current interest to the business community. Our position is to meet this challenge by advancing three major thrusts:

- Creation and or augmentation of a reference software toolkit with well-defined component interfaces and reference implementations to support integration and interoperation with existing management software and to enable application developers from different compute levels to build upon hardened components to create customizable workflow solutions.

- Community engagement with not only DOE scientists, workflows and facility communities but also developers and stakeholders from the private sectors to harmonize the disparate and stove-piped workflow landscape.

- Research and development activities that will ensure that scientific workflows can be efficiently deployed on not only high-end systems but also future exascale desktop machine and a vast distributed set of cloud resources and meet multi-level computing-specific workflow challenges, such as providing high performance and scalable workflow execution, scheduling heterogeneous workloads on heterogeneous systems, enabling flexible and granular communication and coordination, and automatically detecting and recovering from faults when tasks are run across multiple levels in compute hierarchy.

Overall, spinning off these efforts sooner rather than later will spin the benefits, much greater than our investments, back to the DOE to accomplish its national security mission much more efficiently and effectively.