Next Generation Workflow Management Software for Globally Distributed Data Sciences at Exabyte Scale

Submitted by:
Kaushik De (kaushik@uta.edu), University of Texas at Arlington
Alexei Klimentov (alexei.klimentov@cern.ch), Brookhaven National Laboratory
Torre Wenaus (wenaus@gmail.com), Brookhaven National Laboratory

Authors: *F. Barreiro, D. Benjamin, M. Borodin, K. De, A. Forti, J. Elmsheuser, L. Heinrich, A. Klimentov, FH. Lin, T. Maeno, M. Muskinja, P. Nilsson, T. Wenaus*.

The ATLAS collaboration at the Large Hadron Collider (LHC) has developed and demonstrated through decades-long successful operation the current paradigm of computing for large experiments. Using the PanDA software ecosystem, thousands of physicists coherently use hundreds of distributed computing centers worldwide as a single virtualized resource. All experimental workflows are executed in parallel through Panda at globally distributed clusters following required priorities and shares.

PanDA manages the concurrent execution of computational work for thousands of users by carefully balancing CPU, disk and tape resources worldwide. Numerous subsystems and external software systems are managed algorithmically. The system has evolved to incorporate the latest innovations in information technology. We propose to evolve PanDA for the next decade of operation.

The past decade of PanDA operations primarily involved homogeneous grid clusters running on x86 CPU farms alongside disk-based storage systems. PanDA is evolving to incorporate heterogeneous technologies, from Kubernetes to edge services for HPCs, architectures incorporating GPUs, TPUs and FPGAs, and hierarchical storage from solid state devices to tape systems.

While networking is at the center of distributed systems, PanDA still awaits the emerging technologies needed to fully control and optimize workflows over the network. New techniques like streaming and caching are being incorporated for some workflows currently - a global optimization of networking is expected in the next generation of PanDA.

Monitoring and load balancing systems are expected to become less algorithmic, developing higher levels of optimization through machine learning. Front end systems like Jupyter notebooks require the power of PanDA to exploit distributed execution for thousands of simultaneous scientific users. Shared resources will be optimized through advanced AI

techniques. Complex workflows in the age of ML will be enabled on massively distributed datasets.

PanDA is designed to manage both highly structured computational tasks as well as arbitrary workflows required by individual physicists on the same distributed resources while accessing massive datasets. New and complex workflows make this orchestration challenging in the future. PanDA is expected to scale up using new software engineering tools, distributed agents and real time artificial intelligence.

The next generation of scientific experiments with huge distributed datasets, for example the HL-LHC and sky surveys, will require new innovations. The PanDA approach will evolve from custom designed experimental solutions for a few experiments to multi-domain usage through advanced software engineering. Hierarchical storage systems with intelligent data balancing based on usage, streaming services, and application managed networking systems will become incorporated in PanDA. We plan for a decade-long period of transformation into the next generation workload system, which will meet and exceed all future requirements in High Energy Physics, and provide an exemplar for many other scientific domains.