Running data-intensive non-HEP scientific applications with PanDA WMS at OLCF

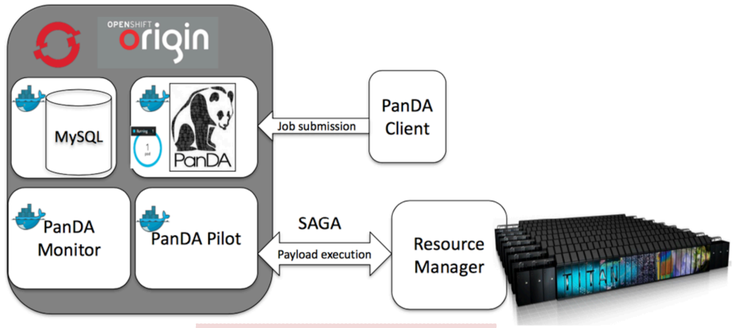
V.G. Anantharaj, F. Barreiro Megino, K. De, S. Jha, J.C. Kincl, A. Klimentov, T. Maeno, R. Mashinistov,
P. Nilsson, D. Oleynik, S. Oral, S. Padolski, S. Panitkin, P. Svirin, M. Turilli, J.C. Wells and T. Wenaus

BigPanDA project goals and highlights

- PanDA Workload Management System (WMS) is being developed by the ATLAS experiment at LHC since 2005, it currently serves ~1400 active users and with ~30 Million jobs per month on hundreds of compute sites around the world
- BigPanDA project: an adaptation of PanDA for HPC as well as use of PanDA for projects and experiments beyond ATLAS
- A DOE ASCR funded project since 2012; a collaboration between BNL, UTA, ORNL and Rutgers since 2015
- See also poster: "BigPanDA Experience on Titan for the ATLAS" Experiment at the LHC"

PanDA infrastructure at OLCF

- PanDA Server at OLCF offers automated tools for shaping, submission, management of jobs, including data transfer
- The PanDA Server instance was installed on OpenShift container management platform
- Users submit their payloads to this instance; in its turn, PanDA Server interacts with edge services on Titan head nodes which execute these payloads



PanDA infrastructure at OLCF

Accomplishment and Future

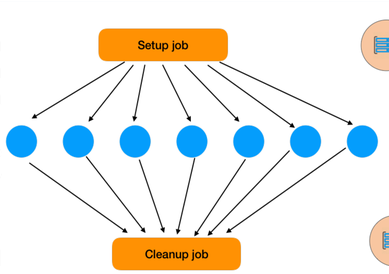
- PanDA infrastructure is installed at OLCF, verified and tested with multiple project queues
- Several projects from various fields of science have run their codes on Titan using automated job submission through PanDA Server at OLCF
- The PanDA Server at OLCF will integrate OLCF and other resources into a distributed computing environment
- More experiments (BELLE2, ALICE, and others) and projects (BlueBrain) have expressed interest to run their payloads on OLCF via PanDA.

Lattice QCD

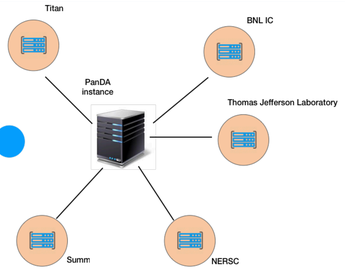
Lattice QCD (LQCD) is a well-established non-perturbative approach to solving the quantum chromodynamics theory of quarks and gluons. Current LQCD payloads can be characterized as massively parallel, occupying thousands of nodes on leadership-class supercomputers. It is understood that future LQCD calculations will require exascale computing capacities and new a workload management system in order to manage throughput efficiently.

Achievements:
- Large LQCD payloads have been successfully tested with PanDA on Titan together with automated data transfer.
- New kinds of payloads will be available for Summit in terms of Early Science Program.

Vision:
- Use PanDA to achieve an integrated approach to all phases of LQCD workloads.



Example of a workflow for LQCD

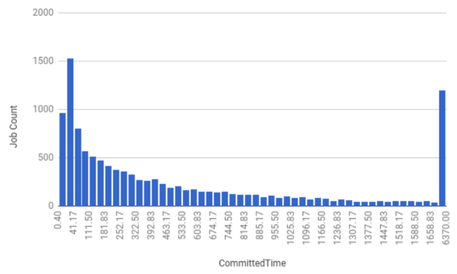


LQCD Future computing infrastructure

USQCD

IceCube

IceCube is the world's largest neutrino detector, encompassing a cubic kilometer of ice at the South Pole. It searches for neutrinos from the most violent astrophysical sources: exploding stars, gamma-ray bursts, and cataclysmic phenomena involving black holes and neutron stars. The IceCube telescope is a powerful tool to search for dark matter and could reveal the physical processes associated with the enigmatic origin of the highest energy particles in nature. IceCube studies the neutrinos at energies that far exceed those produced by accelerator beams. In order to understand experimental details of neutrino-event observations, large-scale detector simulations are needed.



Walltime distribution for IceCube payloads on Titan

Achievements:
- IceCube payloads were tested in Singularity containers on Titan using job shaping.
- Input and output data exchange via GridFTP tested for IceCube jobs on Titan.

Next step:
- Execute a campaign with 35,000 files to be processed on Titan in a backfill mode.

Vision:
- Integrate Titan into IceCube's production workflow.

IceCube South Pole Neutrino Observatory

Molecular Dynamics

Current development efforts of the project are focused on advancing protein simulations where chemistry is linked with large-scale protein conformational dynamics and free energy computations. The application involves, but not limited to, kinases, ATP hydrolases, and DNA repair enzymes. These biomolecules present a unique set of challenges that present experimental techniques have difficulty addressing, while theory can provide detailed understanding at the atomic level.

Achievements:
- CHARMM payload (hybrid MPI/OpenMP/GPU) example built and executed on Titan
- Depending on the type of projects, payloads can expand beyond 500 nodes on Titan; currently, it uses 60-124 nodes for each project

PMI

Computational Biology

Biosciences Division at ORNL is focused on understanding complex biological systems and their relationship with the environment. It is oriented toward aspects of genome and proteome analysis, and molecular systems biology. 2D genome scans using GBOOST software have been performed using PanDA. These highly parallel payloads enable one to understand the organization of genetic regulatory networks. Quantitative genetic analyses are employed that identify combinations of genetic loci that explain traits of economic importance, e.g., such as human health.

Achievements:
- GBOOST payload (GPU) example built and tested on Titan with PanDA

cbi THE CENTER FOR BIOENERGY INNOVATION

nEDM

Precision measurements of the properties of the neutron present an opportunity to search for violations of fundamental symmetries and to make critical tests of the validity of the Standard Model of electroweak interactions. The goal of the nEDM experiment at the Fundamental Neutron Physics Beamline at the Spallation Neutron Source (ORNL) is to further improve the precision of this measurement by a factor of 100. nEDM experiment requires detailed simulation of the detector.

Achievements:
- Detailed nEDM detector simulations were executed on Titan via PanDA Server at OLCF
- Currently nEDM prepares for a future computational campaign

Vision:
- Use PanDA Server at OLCF to execute nEDM data challenge and, their production workload.

nEDM

LSST/DESC

A goal of LSST (Large Synoptic Survey Telescope) project is to conduct a 10-year survey of the sky that is expected to deliver 200 petabytes of data after it begins full science operations in 2022. The project will address some of the most pressing questions about the structure and evolution of the universe and the objects in it. It will require a large amount of simulations, which model the atmosphere, optics and camera to understand the collected data.

Achievements:
- Phosim simulations were run on Titan using PanDA Server at OLCF
- Phosim long-running jobs required exploration of checkpointing capabilities on Titan

Vision:
- Evaluate for production WLMS

LSST Large Synoptic Survey Telescope

The University of Texas ARLINGTON · BROOKHAVEN NATIONAL LABORATORY · OAK RIDGE National Laboratory · RUTGERS UNIVERSITY · U.S. DEPARTMENT OF ENERGY