PanDA WMS experience with OLCF beyond ATLAS

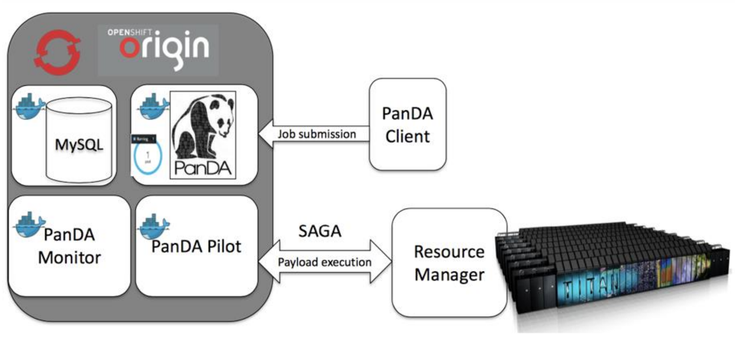
V.G. Anantharaj, F. Barreiro Megino, K. De, S. Jha, J.C. Kincl, A. Klimentov, T. Maeno, R. Mashinistov, P. Nilsson, D. Oleynik, S. Oral, S. Padolski, S. Panitkin, P. Svirin, M. Turilli, J.C. Wells and T. Wenaus

PanDA/OLCF project goals and highlights

- ❑ PanDA Workload Management System (WMS) is developed for ATLAS experiment at LHC since 2005, currently serves ~1400 active users and with ~30 mln jobs per month
- ❑ BigPanDA project: an adaptation of PanDA for HPC and other projects and experiments, a DOE ASCR funded project since 2012
- ❑ The goal of the collaboration of BigPanDA and OLCF: Translate the R&D artifacts and accomplishments from the BigPanDA into LCF operational advances and enhancement

PanDA infrastructure at OLCF

- ❑ PanDA infrastructure at OLCF offers automated job submission to its users
- ❑ A dedicated instance of PanDA Server was installed on OpenShift container management platform at OLCF
- ❑ This PanDA Server instance now interacts with edge services which submit payloads to Titan and manage their execution



PanDA infrastructure at OLCF

Accomplishment and Future

- ❑ PanDA infrastructure has been successfully installed and tested at OLCF, working, tested with multiple project queues
- ❑ Several projects from various fields of science ran their codes on Titan using automated job submission through PanDA Server at OLCF
- ❑ In the future it 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.

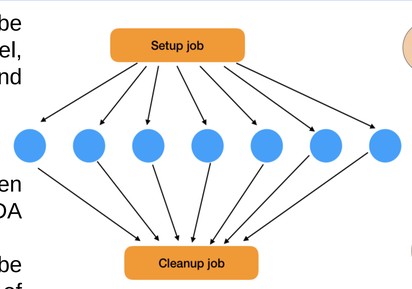
Lattice QCD

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 workload management system in order to manage them efficiently.

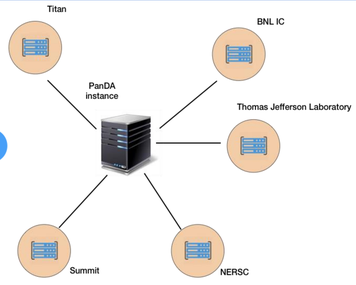
LQCD payloads on Titan can be described as massively parallel, occupying thousands of nodes and GPUs, occupying 10-20 hours.

Achievements:

- ❑ Large LQCD payloads have been successfully tested with PanDA on Titan
- ❑ New kinds of payloads will be available for Summit in terms of Early Science Program



Example of a workflow for LQCD

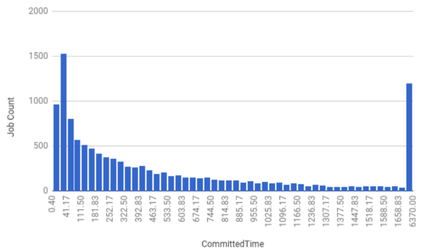


Future computing infrastructure for LQCD

USQCD

IceCube

IceCube is the world's largest neutrino detector, encompassing a cubic kilometer of ice at the South Pole. IceCube searches for neutrinos from the most violent astrophysical sources: events like 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. In addition, exploring the background of neutrinos produced in the atmosphere, IceCube studies the neutrinos themselves; their energies far exceed those produced by accelerator beams. In order to understand experimental details of neutrino events observations large-scale detector simulations are needed.



Walltime distribution for IceCube payloads on Titan

Achievements:

- ❑ Input and output data exchange via GridFTP was tested for IceCube jobs on Titan
- ❑ Realistic payloads were tested in Singularity containers on Titan

Next step:

- ❑ run a campaign with 35000 files to be processed on Titan in a backfill mode

IceCube South Pole Neutrino Observatory

Molecular Dynamics

Current developmental efforts of the project are focused on advancing the ability to simulate protein systems where chemistry is linked with large-scale protein conformational dynamics and corresponding free energy computations. The application part involves, but not limited to, kinases, ATP hydrolases and DNA repair enzymes. These biomolecules present unique set of challenges that present experimental techniques have difficulties to address, 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

The Biosciences Division at ORNL is focused on advancing science and technology to better understand complex biological systems and their relationship with the environment. It is oriented toward various aspects of genome and proteome analysis, and molecular systems biology and spans multiple domains of science. 2D genome scans using GBOOST software have been performed using PanDA. These highly parallel payloads allow to capture the organization of genetic regulatory networks. Quantitative genetic analyses are employed that identify the combinations of genetic loci that explain the trait. These analyses will enable characterization of the genetic networks underlying traits of economic importance, 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 another 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

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

LSST Large Synoptic Survey Telescope