

# Next Generation Workload Management for Pre-Exascale Science: BigPanDA

P. Svirin, S. Wilkinson, J. Kincl, R. Mashinistov, A. Merzky, D. Oleynik, S. Oral, S. Panitkin, M. Turilli, K. De, S. Jha, A. Klimentov, J. Wells, T. Wenaus



## Project Vision

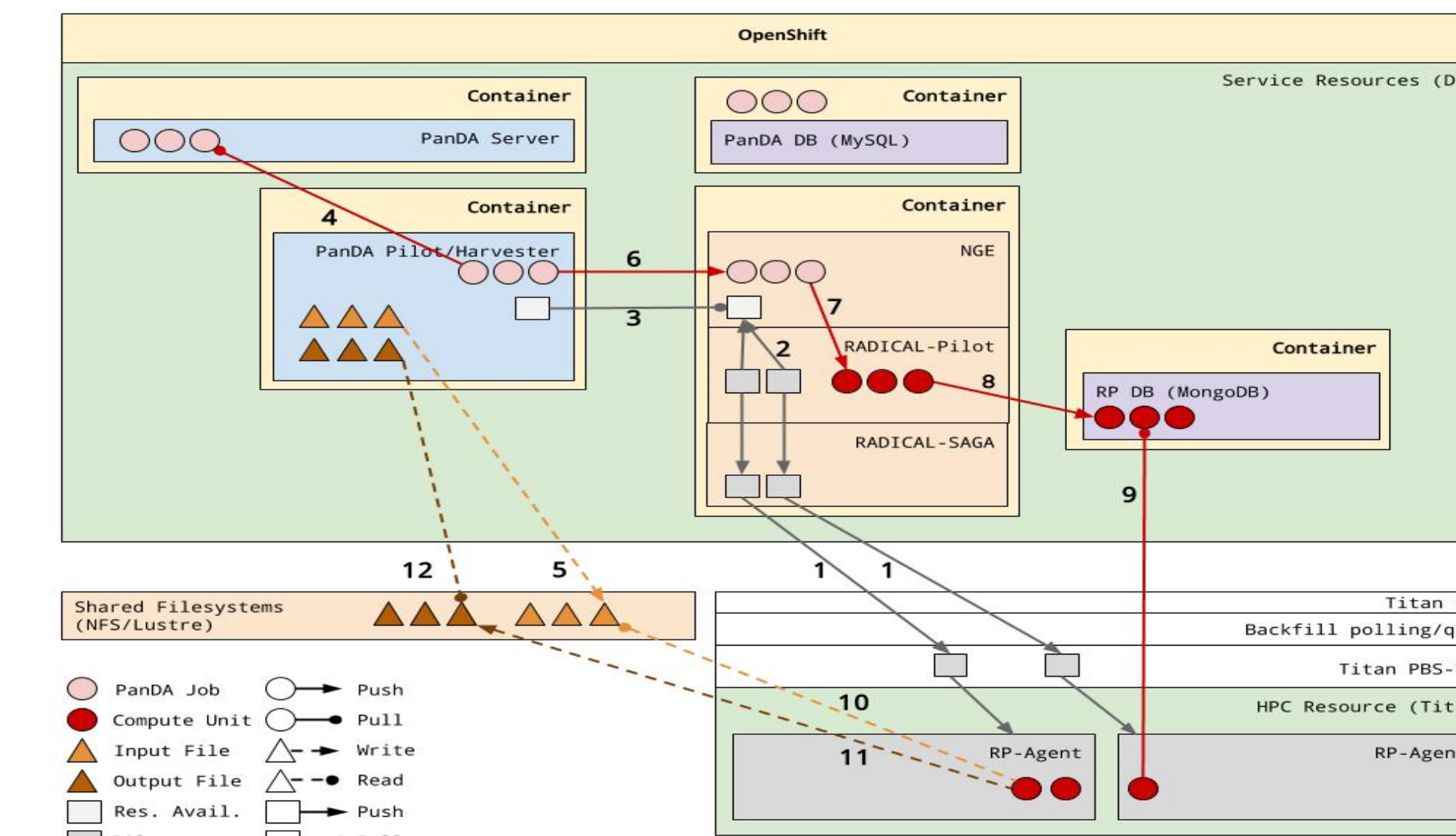
- Translate the R&D artifacts and accomplishments from the BigPanDA and AIMES projects into LCF operational advances and enhancement.
- Utilize compute cycles that would be otherwise unusable and the ability to increase the overall utilization on Titan.
- Extend and generalize the ability of using HPC for HTC for ATLAS and other communities.

## Accomplishment, Impact, and Future

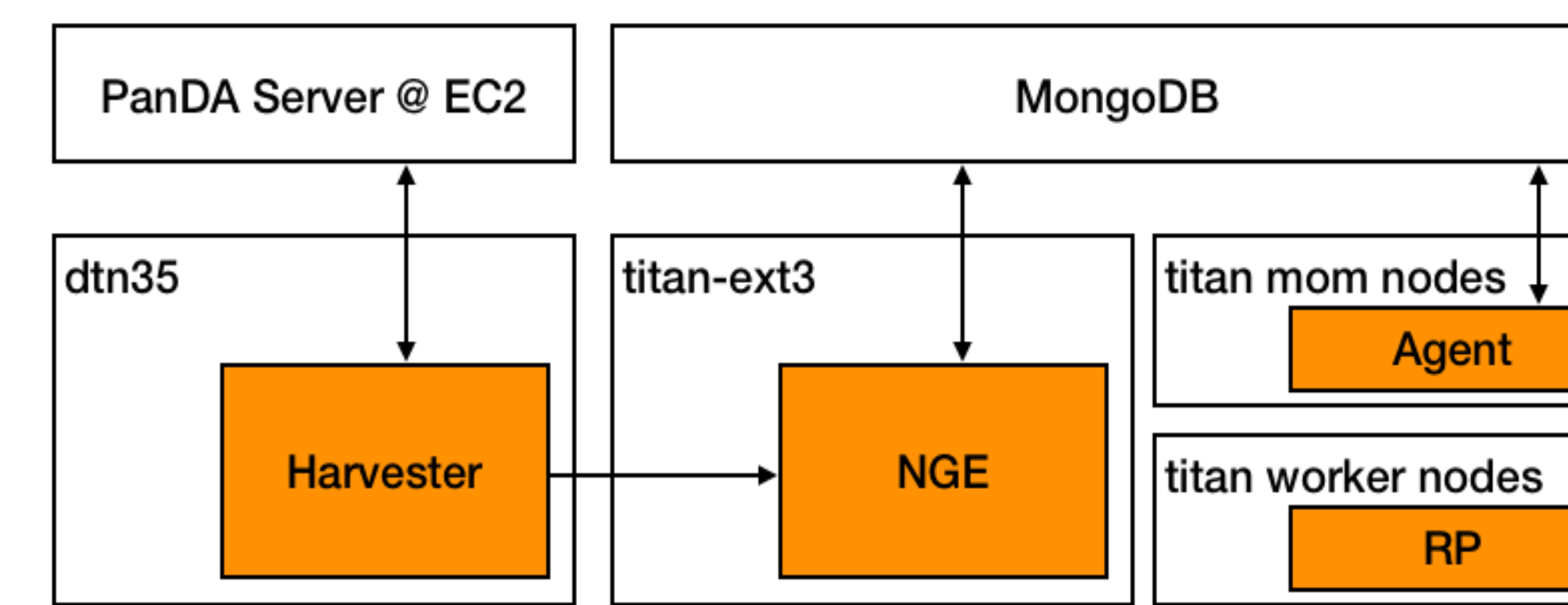
- OLCF resources are utilized using the latest PanDA software versions.
- Optimized execution for payloads on Titan using Next-Generation Executor (NGE). New payload types to be tested with Harvester/NGE integration.
- Continue to develop current software to support heterogeneous resources and diverse workflows for multiple experiments and projects beyond ATLAS.
- Web site: <http://news.pandawms.org/bigpanda.html>
- GitHub: <https://github.com/PanDAWMS>

## Adaptive Execution

- Next-Generation Executor(NGE) conforms to model of Pilot Abstraction: <https://arxiv.org/abs/1508.04180>
- Flexible Execution: No assumption about number of events per simulation. Better utilization via multiple execution generations
- Investigate the advantages of scale, flexible execution and interoperability outside production constraints on HPC resources.
- Enables PanDA to be independent from: Workload heterogeneity:
  - No need of tailoring MPI scripts to the specifics of each type of workload.
  - Resource heterogeneity: No need of implementing tailored support for each HPC machine flavour
- Modeling probability of allocation utilization (ALCC) as a function of execution strategy.
- Supporting backfill/regular queueing on Titan; concurrent execution of workloads on the same pilot; distribution of jobs on Titan and Summit supercomputers.
- Future plans:
  - Extending profiling and analytics capabilities of NGE stack to Harvester stack executions.
  - Port design features to Yoda.

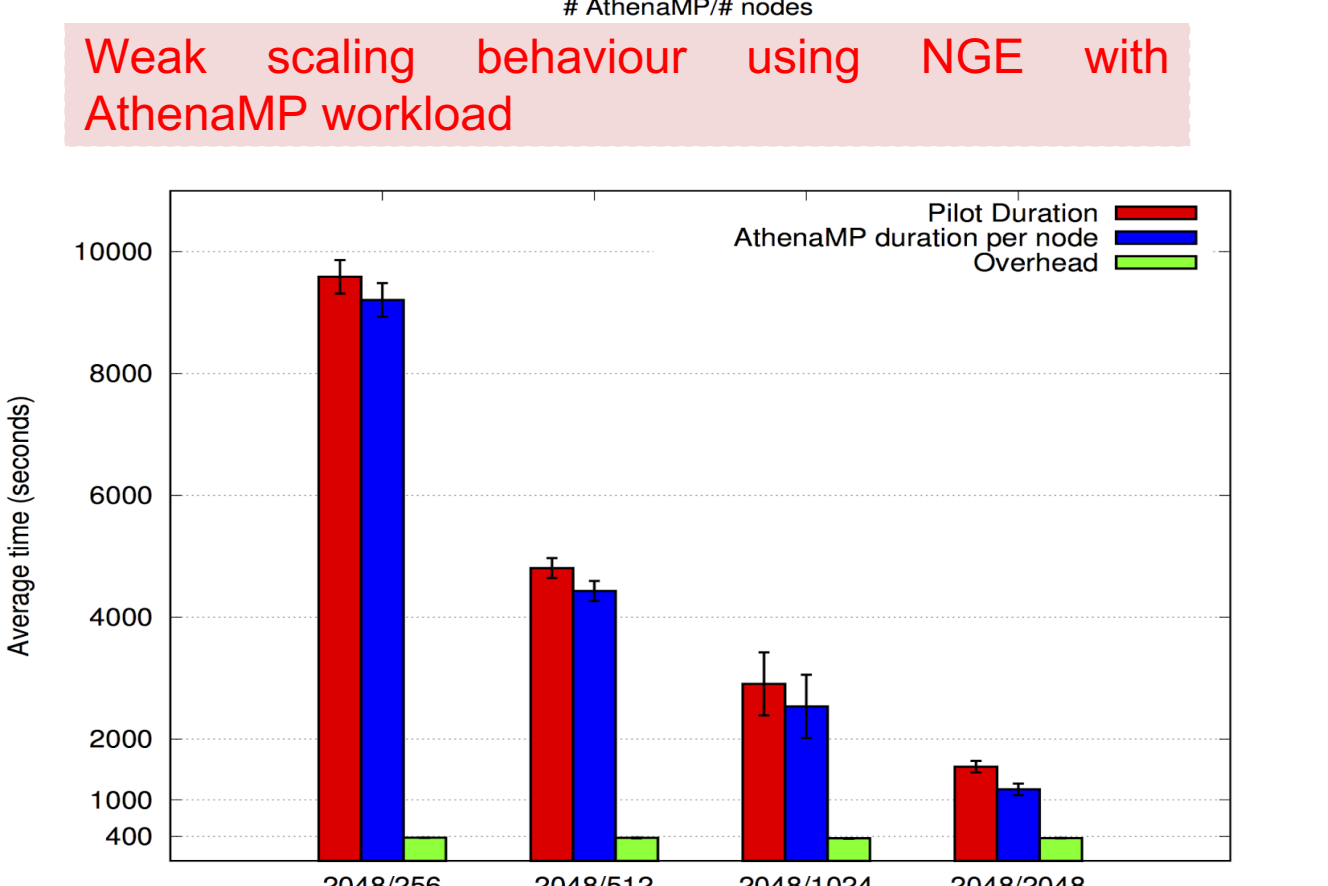
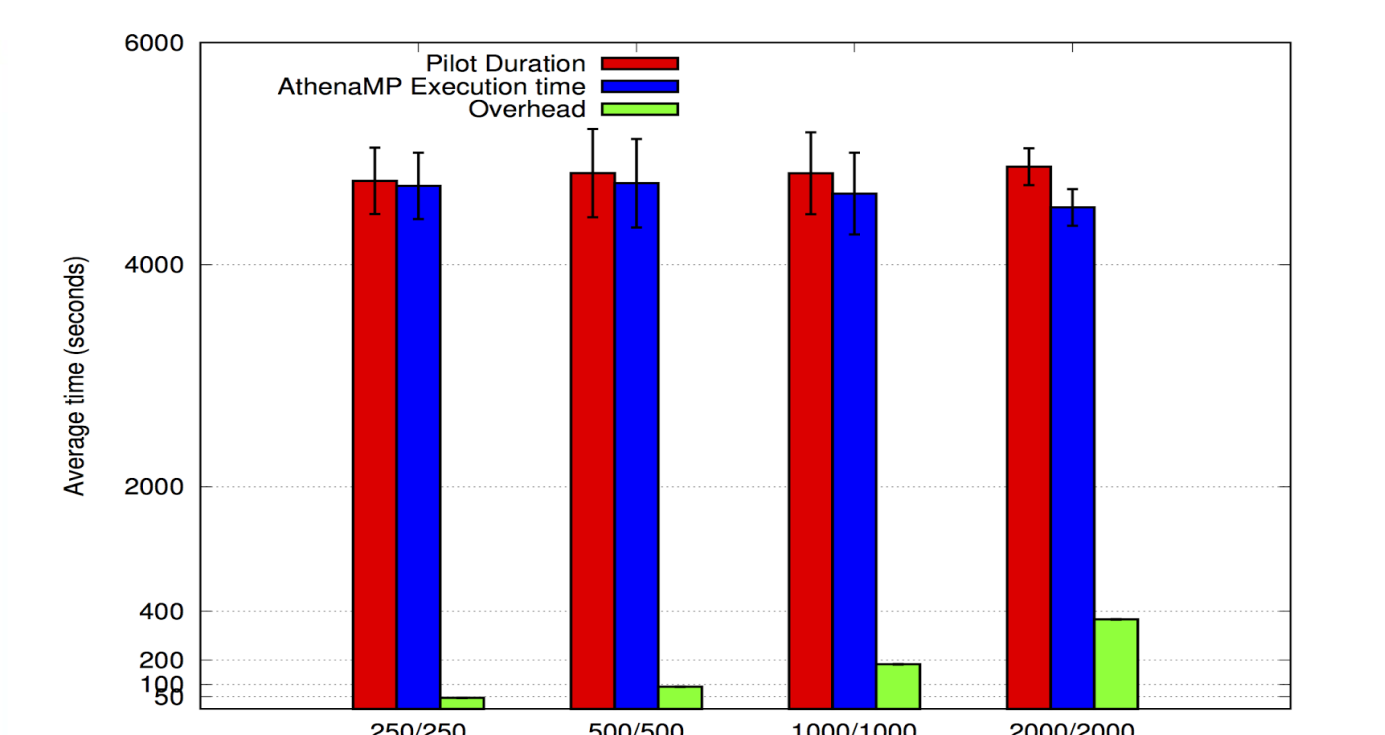


PanDA-NGE Integration: Architecture

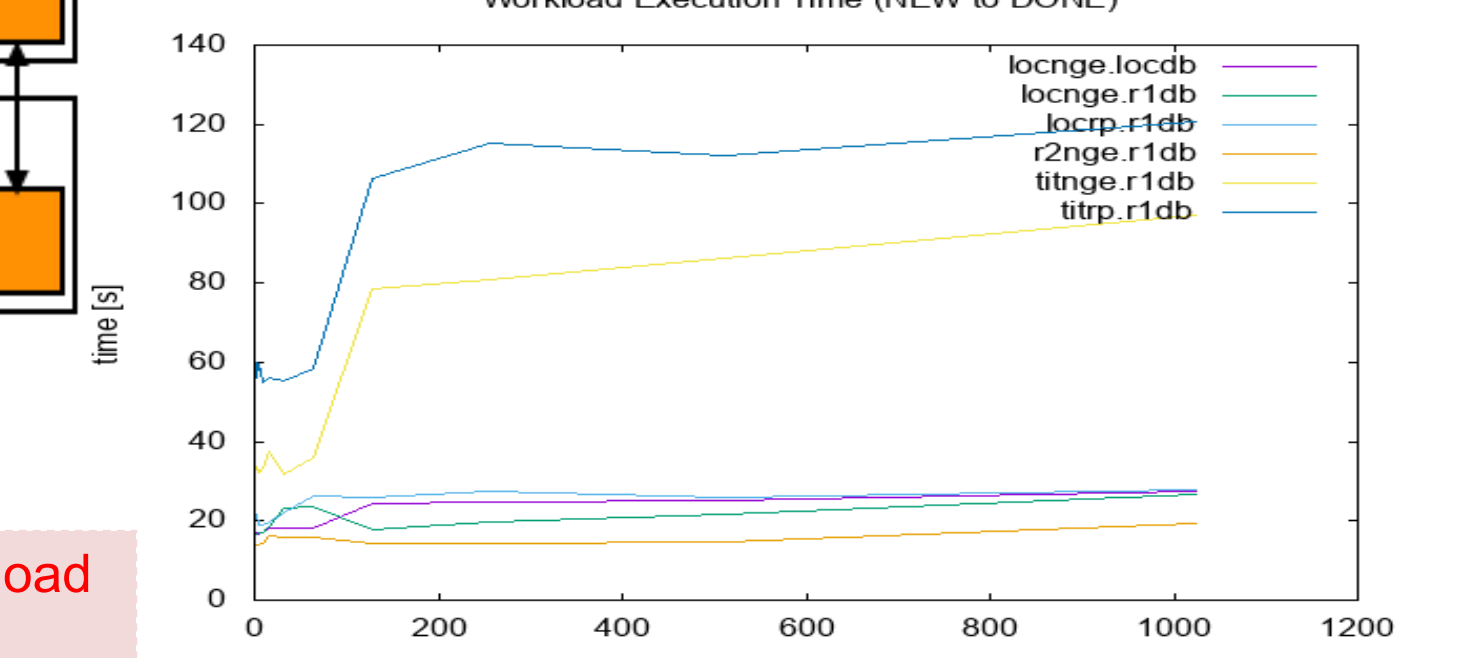


PanDA-NGE Integration: deployment on Titan

NGE performance : null workload execution

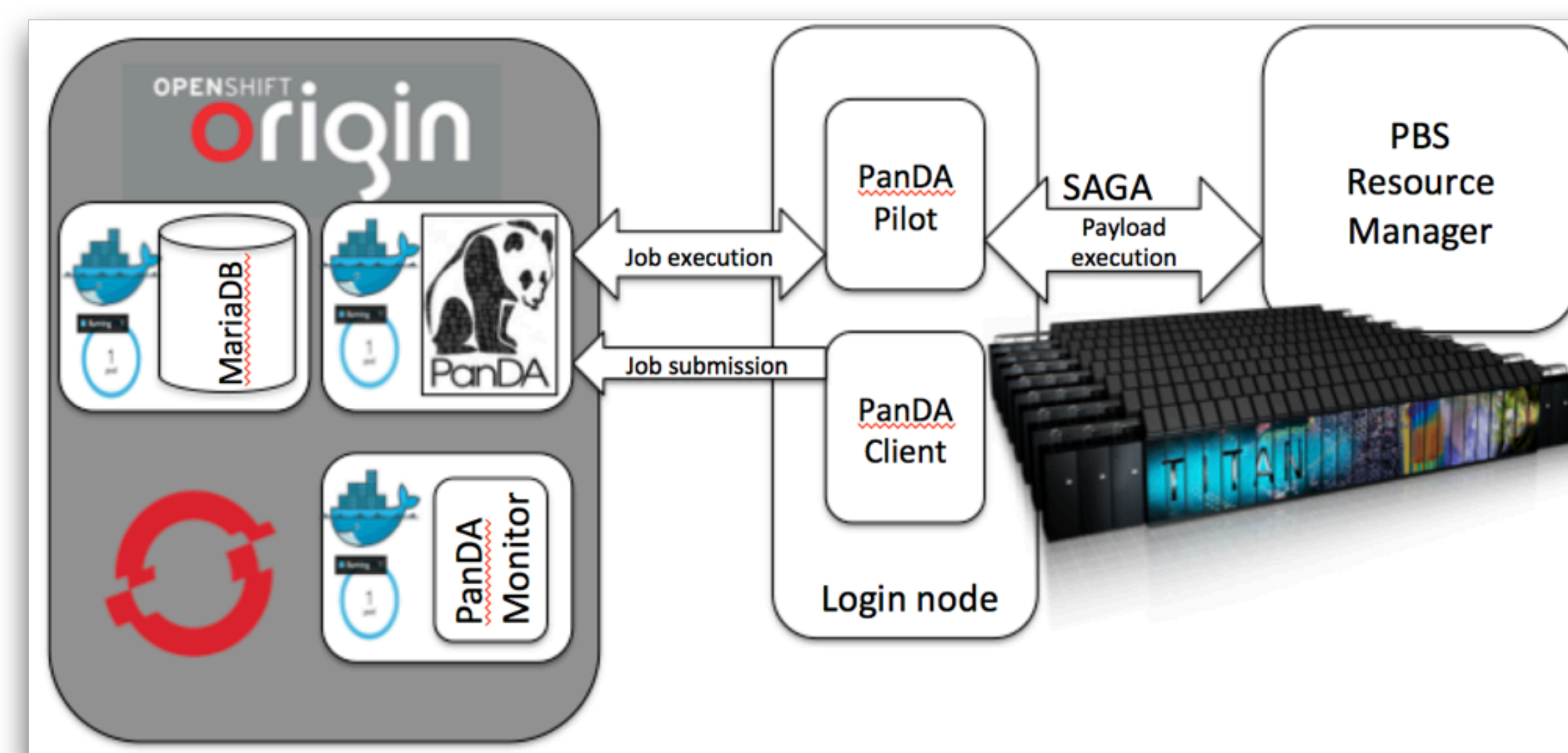


Strong scaling behaviour for NGE with AthenaMP workload using 100 events per AthenaMP task.



## Operations

- A VO-independent PanDA Server has been set up in Amazon EC2. It is used for non-ATLAS projects that utilise heterogeneous Grid and HPC resources.
- In March 2017 a new PanDA server instance has been set up at OLCF to serve various experiments.
- A set of demonstrations serving diverse scientific workflows including particle physics experiments, biology studies of the genes and human brain, and molecular dynamics studies was implemented.



PanDA Server instance running at OLCF

Experiment	Payload/SW	Number of jobs per campaign	Number of nodes per job	Walltime (min)	Input data size per job	Output data size per job
Genomics	GBOOST	10	2	30	100 M	300 M
Molecular Dynamics	CHARMM	10	124	30-90	10 K	2-6 GB
IceCube	NuGen	4500K	1	120	500 K	10 KB - 4 GB
LSST/DESC	Phosim	20	2	600	700 MB	70 MB
LQCD	QDP++	10	8000	700	40 GB	150MB
nEDM	GEANT	10	200	20	120 MB	20 MB

Characteristics for non-ATLAS payloads tested with PanDA on Titan

## USQCD

- Lattice quantum chromodynamics (LQCD) is the lattice discretized theory of the strong nuclear force
- A distributed environment for LQCD computations has been set up using PanDA Server instance deployed at the Amazon Cloud
- Variety of payloads, MPI and non-MPI, GPU-based, workflows and independent jobs
- Production campaign has been started on BNL Institutional Cluster
- Data transfers between the sites have been tested with Globus Online

## PMI

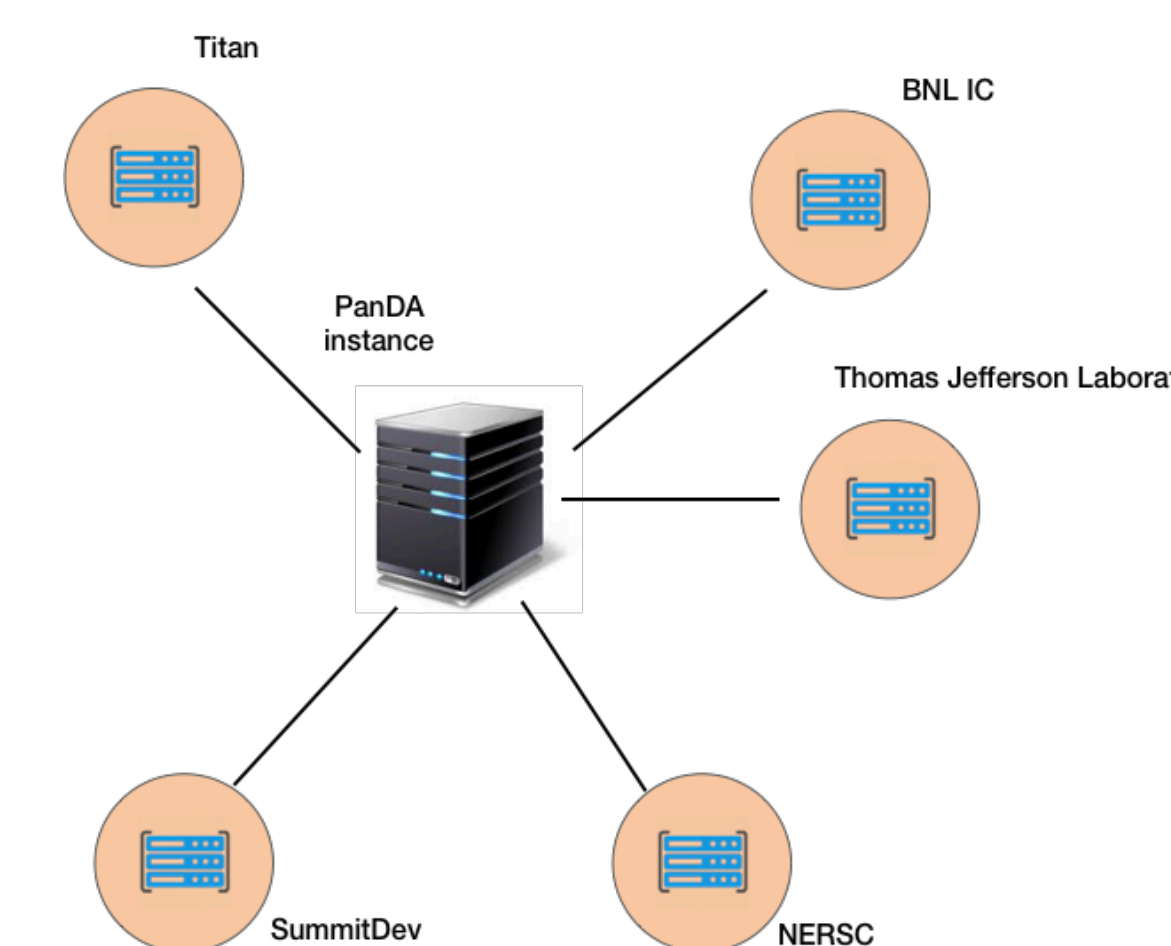
- Simulating Enzyme Catalysis, Conformational Change, and Ligand Binding/Release.
- CHARMM payload (hybrid MPI/OpenMP/GPU) payloads were tested with PanDA on Titan
- Depending on the type of projects, payloads can expand beyond 500 nodes on Titan; currently, it uses 60-124 nodes for each project
- Payloads have been also tested with Harvester/NGE integration

## nEDM

- The goal of the nEDM (Neutron Electric Dipole Moment) experiment to further improve the precision of neutron properties measurement by a factor of 100
- Detailed nEDM detector simulations were executed on Titan via PanDA WMS
- Currently nEDM prepares for a future computational campaign
- Short time of processing of events makes nEDM good candidate for backfill consumption at OLCF

## LSST

- 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
- LSST simulations were run on Titan using PanDA Server at OLCF
- Phosim long-running jobs required exploration of checkpointing capabilities on Titan
- Grid environment with 36 endpoints for LSST was configured and tested with PanDA Server in Amazon Cloud



Future computing infrastructure for LQCD

- New kinds of payloads will be available for Summit in terms of Early Science Program



- In collaboration with Center for Bioenergy Innovation at ORNL, the PanDA based workflow for epistasis research was established. Epistasis is the phenomenon where the effect of one gene is dependent on the presence of one or more 'modifier genes'

- The GBOOST application, a GPU-based tool for detecting gene-gene interactions in genome-wide case control studies, was tested on Titan with PanDA



- The IceCube Neutrino Observatory is the first detector of its kind, designed to observe the cosmos from deep within the South Pole ice.
- IceCube payloads were tested in Singularity containers on Titan using job shaping.
- PanDA allows to combine IceCube's non-MPI payloads into assemblies and executes them as a simple MPI application
- Input and output data exchange via GridFTP tested for IceCube jobs