PanDA WMS for Lattice QCD Computations

S. Panitkin, R. Edwards, R. Larsen, S. Mukherjee, P. Svirin, D. Trewartha

US Lattice Quantum Chromodynamics

Lattice quantum chromodynamics (LQCD) is the lattice discretized theory of the strong nuclear force, the force that binds quarks together into particles such as the proton and neutron. High precision predictions from LQCD are required for testing the standard model of particle physics, a task with increased importance in the era of the Large Hadron Collider (LHC), where deviations between numerical LQCD predictions and experiment could be signs of new physics. LQCD also has a vital role to play in nuclear physics, where such calculations are used to compute and classify the excited states of protons, neutrons and other hadrons; to study hadronic structure; and to compute the forces and binding energies in light nuclei.

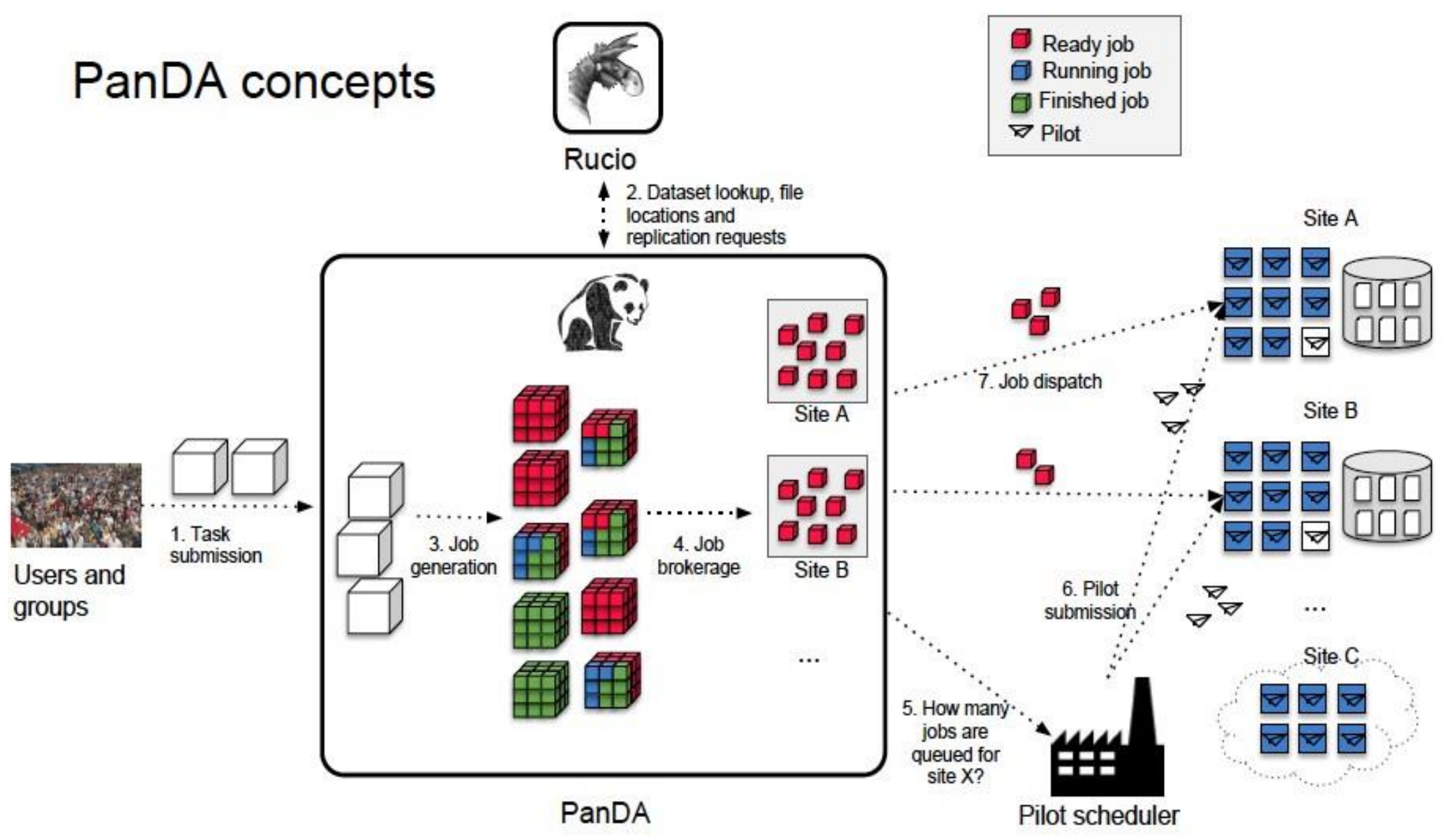
LQCD is a grand challenge subject, with large-scale computations consuming a considerable fraction of publicly available supercomputing resources. The computations typically proceed in two phases: in the first phase, one generates thousands of configurations of the strong force fields (gluons), colloquially referred to as gauge fields. This computation is a long-chain Monte Carlo process, requiring the focused power of leadership class computing facilities for extended periods. In the second phase, these configurations are analyzed. Until a few years ago, the analysis phase would often account for a relatively small part of the cost of the overall calculation. In recent years, however, focus has turned to more challenging physical observables and new analysis techniques that demand solutions often require an equal or greater amount of computation than gauge field generation.

It is understood that future LQCD calculations will require exascale computing capacities and that a robust workload management system (WMS) is necessary in order to manage them efficiently. PanDA WMS is a good choice for workload and data transfer management at this scale.

PanDA WMS

PanDA WMS was developed for the ATLAS Experiment at LHC for job scheduling on the distributed computational infrastructure. The system has been designed to meet ATLAS production and analysis requirements for a data-driven workload management system capable of operating at LHC data processing scale. Currently, as of 2018, PanDA WMS manages processing of over one million jobs per day, serving thousands of ATLAS users worldwide. It is capable of executing jobs on heterogeneous distributed resources which include WLCG, supercomputers, and public and private clouds.

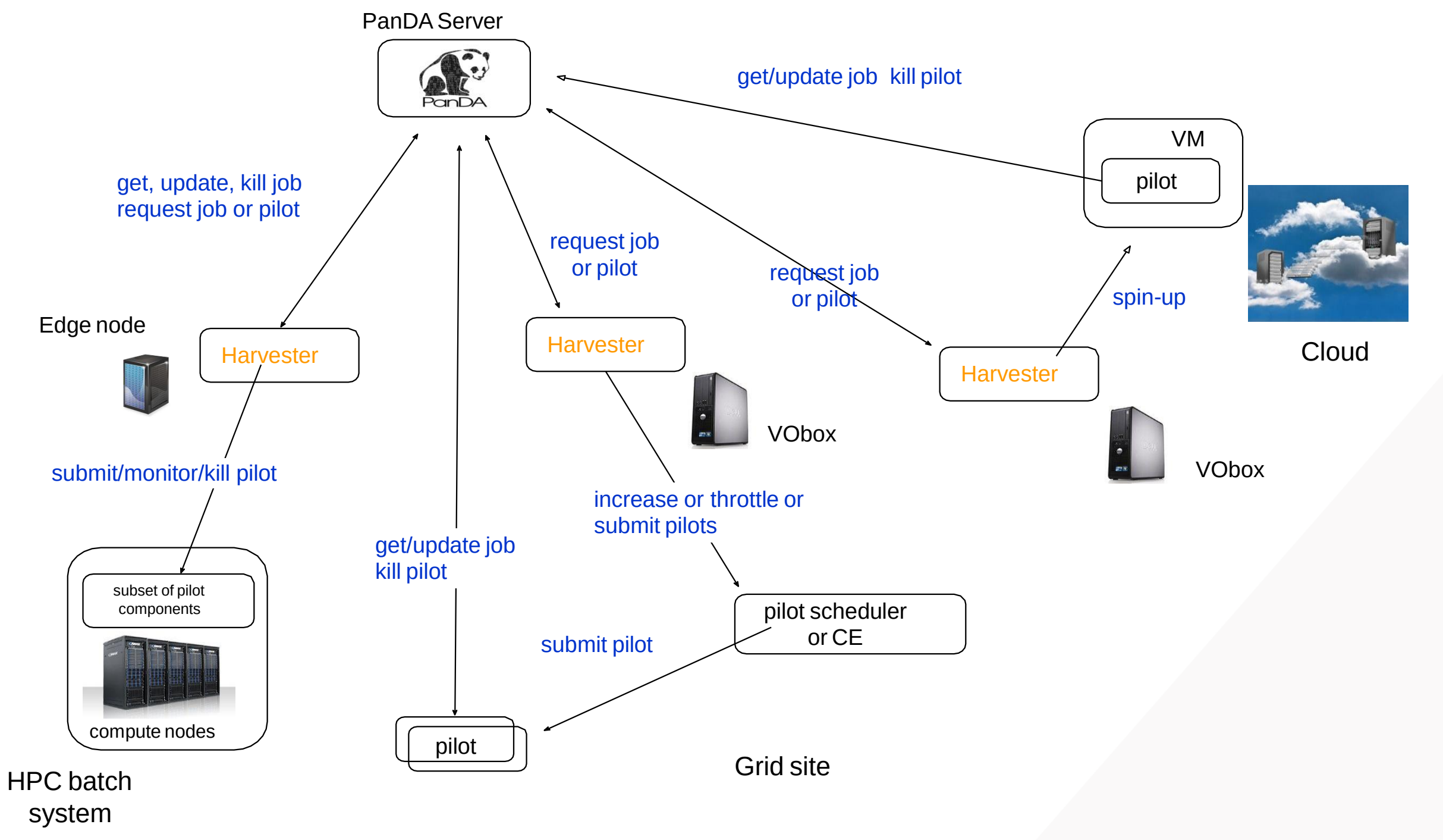
In 2017 PanDA was selected as a WMS for SciDAC-4 supported LQCD project.



- Modular, scalable and extensible design
- Single point of job submission for users
- Can integrate heterogeneous computing resources (Grids, HPCs, clouds, BOINC, …)
- Pilots/agents for job execution
- Secure – X509 authentication, HTTPS communication
- Integrated data transfer capabilities
- Integrated job/task monitoring
- Proven 10+ year record in ATLAS, providing resources on hundreds of sites to thousands of users
- Big Data WMS - since 2013 more than Exabyte of data is being processed every year in ATLAS with PanDA
- Adopted by many experiments and projects: ATLAS, AMS, COMPASS, LQCD, NICA. Tested for DESC/LSST, IceCube, nEDM, BlueBrain, biology, paleo genomics and molecular dynamics workflows

Harvester

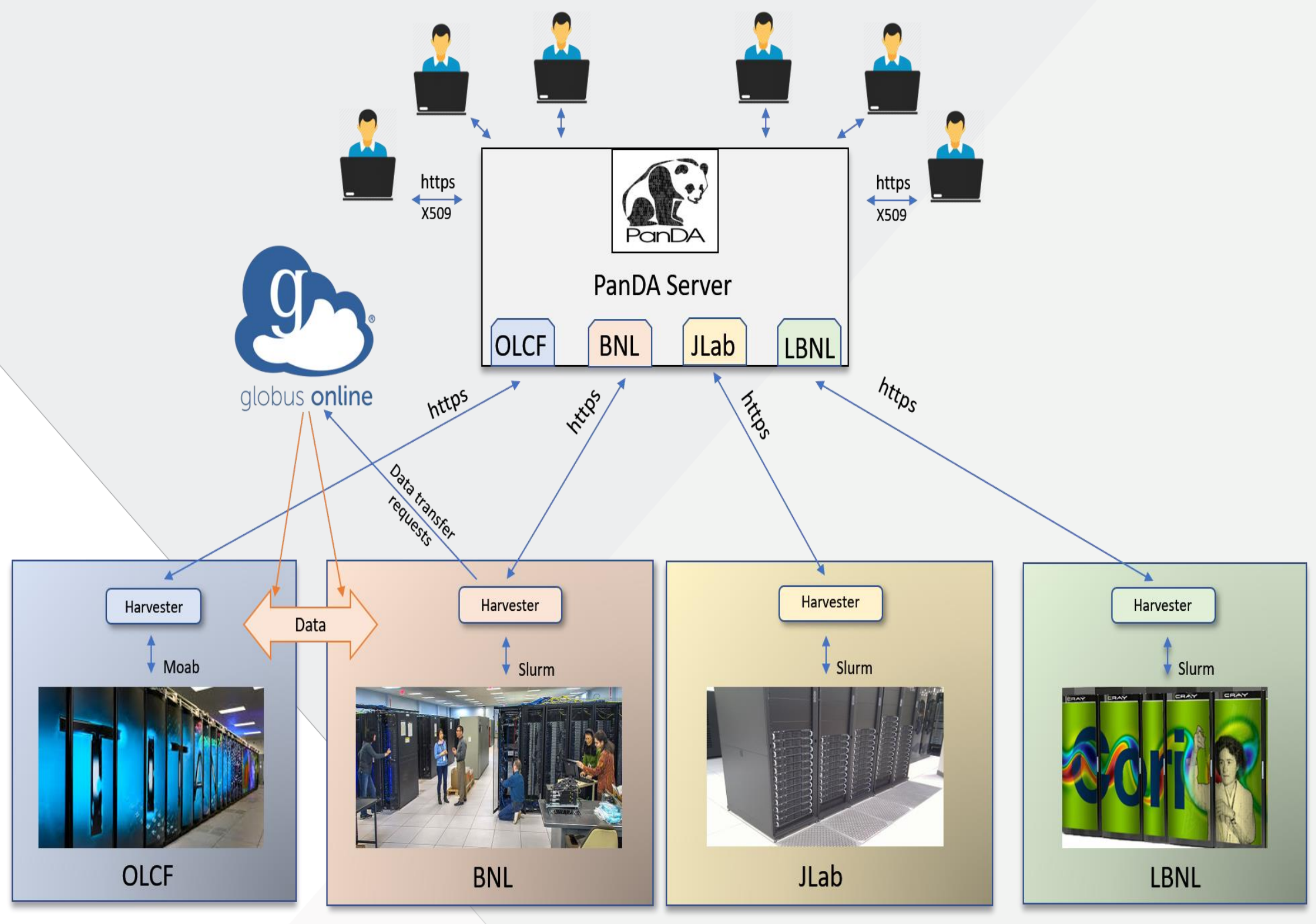
Harvester is a next-generation resource-facing service between PanDA server and collection of workers. Worker is a generalization of a pilot concept and can be, depending on a resource and workflow, a pilot, an MPI job, or a virtual machine. Harvester has modular multithreaded design to support heterogeneous resources and to accommodate for special workflows requirements. Harvester provides for flexible scheduling of job execution and asynchronous data transfer to and from the controlled resource.
Currently used for integration of PanDA servers with Grids, supercomputers and clouds.



Typical LQCD workloads

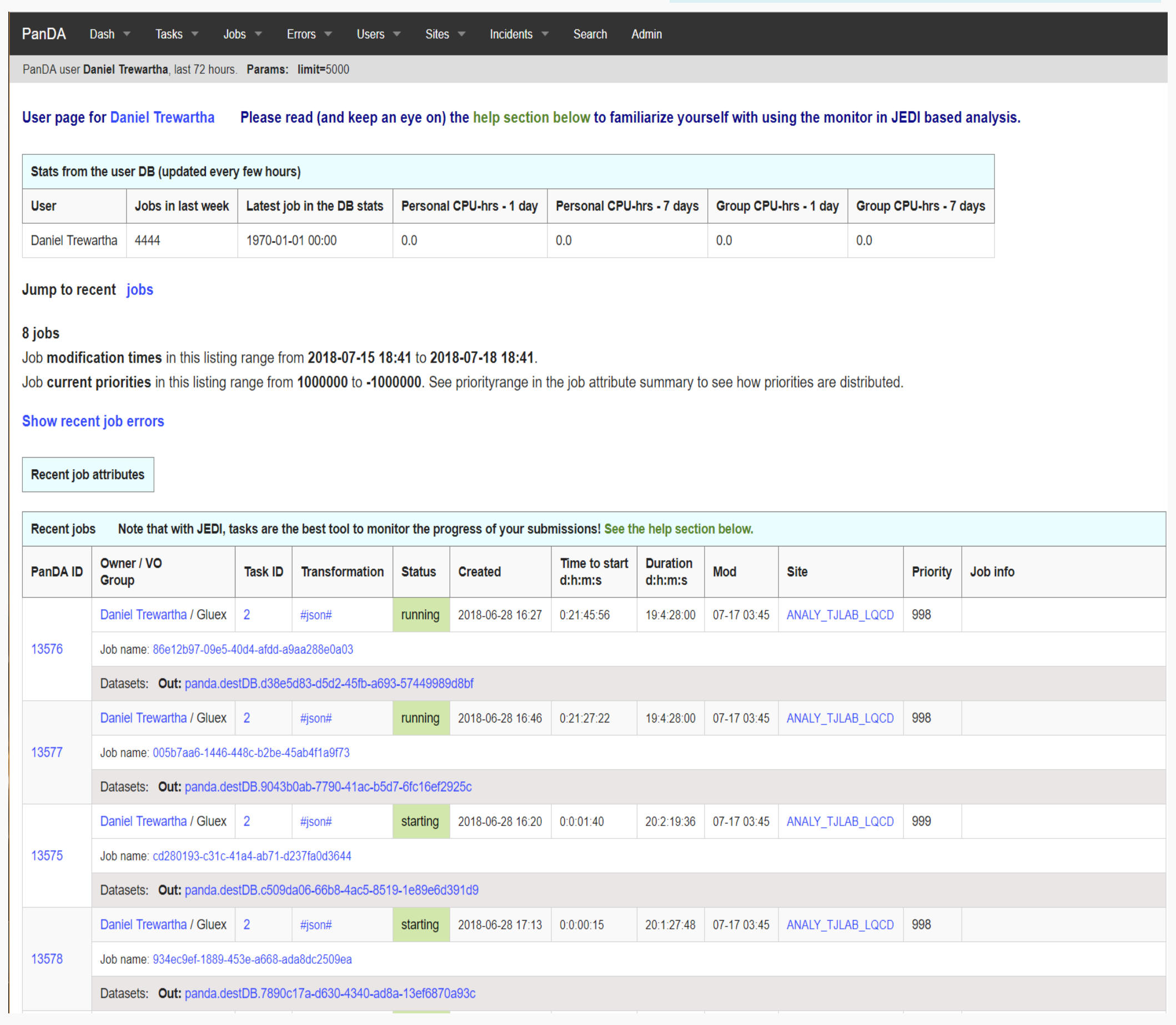
- **Quark line contractions:**
  - –O(10^6) of independent single node jobs, walltime: 2 hours + validation script for jobs output to be run, ~30 minutes, may be run per each job or per bunch
  - –Input data: gauge fields configurations, O(1000) files, ~2GB each
  - –Output: ~100 MB per job
- **Perambulation calculations:**
  - –O(10^6) jobs
  - –MPI jobs, linear scaling with cores change (cores count must be a multiple of 16)
- **QGP calculations:**
  - –input data: 255 configurations ~13 TB
  - –output data so far: ~176 GB
  - –255 configurations*6 sets each = 1330 jobs
  - –job walltime: ~9-12 hours

PanDA setup for LQCD



- SciDAC-4 supported project
- Collaboration between LQCD physicist from BNL and JLab and PanDA development team
- PanDA server on Amazon EC2
- Harvester used for integration with local resources
- Currently Titan at OLCF, Cori at NERSC, LQCD cluster at JLab and Institutional Cluster (IC) at BNL are integrated
- Harvester on Summitdev was tested, Summit supercomputer at OLCF is planned for integration, as well as other sites used by LQCD SciDAC project
- Data transfers with PanDA for LQCD jobs using Globus Online third party transfers tested
- Tools for LQCD specific sequential workflows developed and tested
- First production campaigns ran on IC at BNL and JLab cluster

PanDA monitor for LQCD jobs



PanDA user **Daniel Trewartha**, last 72 hours. **Params: limit**=5000

User page for Daniel Trewartha   Please read (and keep an eye on) the help section below to familiarize yourself with using the monitor in JEDI based analysis.

Stats from the user DB (updated every few hours)

User	Jobs in last week	Latest job in the DB stats	Personal CPU-hrs - 1 day	Personal CPU-hrs - 7 days	Group CPU-hrs - 1 day	Group CPU-hrs - 7 days
Daniel Trewartha	4444	1970-01-01 00:00	0.0	0.0	0.0	0.0

Jump to recent jobs

8 jobs
Job **modification times** in this listing range from **2018-07-15 18:41** to **2018-07-18 18:41**.
Job **current priorities** in this listing range from **1000000** to **-1000000**. See priorityrange in the job attribute summary to see how priorities are distributed.

Show recent job errors

Recent job attributes

Recent jobs   Note that with JEDI, tasks are the best tool to monitor the progress of your submissions! See the help section below.

PanDA ID	Owner / VO Group	Task ID	Transformation	Status	Created	Time to start d:h:m:s	Duration d:h:m:s	Mod	Site	Priority	Job info
13576	Daniel Trewartha / Gluex	2	#json#	running	2018-06-28 16:27	0:21:45:56	19:4:28:00	07-17 03:45	ANALY_TJLAB_LQCD	998	
	Job name: 86e12b97-09e5-40d4-afdd-a9aa288e0a03										
	Datasets: **Out:** panda.destDB.d38e5d83-d5d2-45fb-a693-57449989d8bf										
13577	Daniel Trewartha / Gluex	2	#json#	running	2018-06-28 16:46	0:21:27:22	19:4:28:00	07-17 03:45	ANALY_TJLAB_LQCD	998	
	Job name: 005b7aa6-1446-448c-b2be-45ab4f1a9f73										
	Datasets: **Out:** panda.destDB.9043b0ab-7790-41ac-b5d7-6fc16ef2925c										
13575	Daniel Trewartha / Gluex	2	#json#	starting	2018-06-28 16:20	0:0:01:40	20:2:19:36	07-17 03:45	ANALY_TJLAB_LQCD	999	
	Job name: cd280193-c31c-41a4-ab71-d237fa0d3644										
	Datasets: **Out:** panda.destDB.c509da06-66b8-4ac5-8519-1e89e6d391d9										
13578	Daniel Trewartha / Gluex	2	#json#	starting	2018-06-28 17:13	0:0:00:15	20:1:27:48	07-17 03:45	ANALY_TJLAB_LQCD	998	
	Job name: 934ec9ef-1889-453e-a668-ada8dc2509ea										
	Datasets: **Out:** panda.destDB.7890c17a-d630-4340-ad8a-13ef6870a93c										