DANILA OLEYNIK
ON BEHALF OF BIGPANDA TEAM

ENABLING BIOLOGY, CHEMISTRY AND OTHER SCIENCES ON TITAN THROUGH BIGPANDA

OUTLINE

- What is PanDA WMS and BigPanDA
  - Success of PanDA for ATLAS Experiment
- PanDA outside ATLAS
  - How PanDA can be used by other communities (dedicated or shared instances)
  - PanDA instance at EC2
    - LSST
    - LQCD
    - nEDM
  - PanDA instance at OLCF
    - IceCube
    - Biology/Genetics
    - Molecular Dynamics
  - PanDA WMS for BlueBrain project

PANDA WORKLOAD MANAGEMENT SYSTEM

- The PanDA workload management system was developed for the ATLAS experiment at the Large Hadron Collider
- A new approach to distributed computing
  - A huge hierarchy of computing centres and opportunistic resources working together
  - Main challenge – how to provide efficient automated performance
  - Auxiliary challenge – make resources easily accessible to all users
- Core ideas :
  - Make hundreds of distributed sites appear as local
    - Provide a central queue for users – similar to local batch systems
  - Reduce site related errors and reduce latency
    - Build a pilot job system – late transfer of user payloads
    - Crucial for distributed infrastructure maintained by local experts
  - Hide middleware while supporting diversity and evolution
    - PanDA interacts with middleware – users see high level workflow
  - Hide variations in infrastructure
    - PanDA presents uniform 'job' slots to user (with minimal sub-types)
    - ***Easy to integrate grid sites, clouds, HPC sites …***
  - Data processing, Production and Analysis users see same PanDA system
    - Same set of distributed resources available to all users
 - Highly flexible – instantaneous control of global priorities by experiment

PANDA AT NUTSHELL

- Pilot based job execution system
  - Pilot manages job execution on local resources, as well as data movement for the job
- Payload is sent only after pilot execution begins on Compute Element
 - Minimize latency, reduce error rates
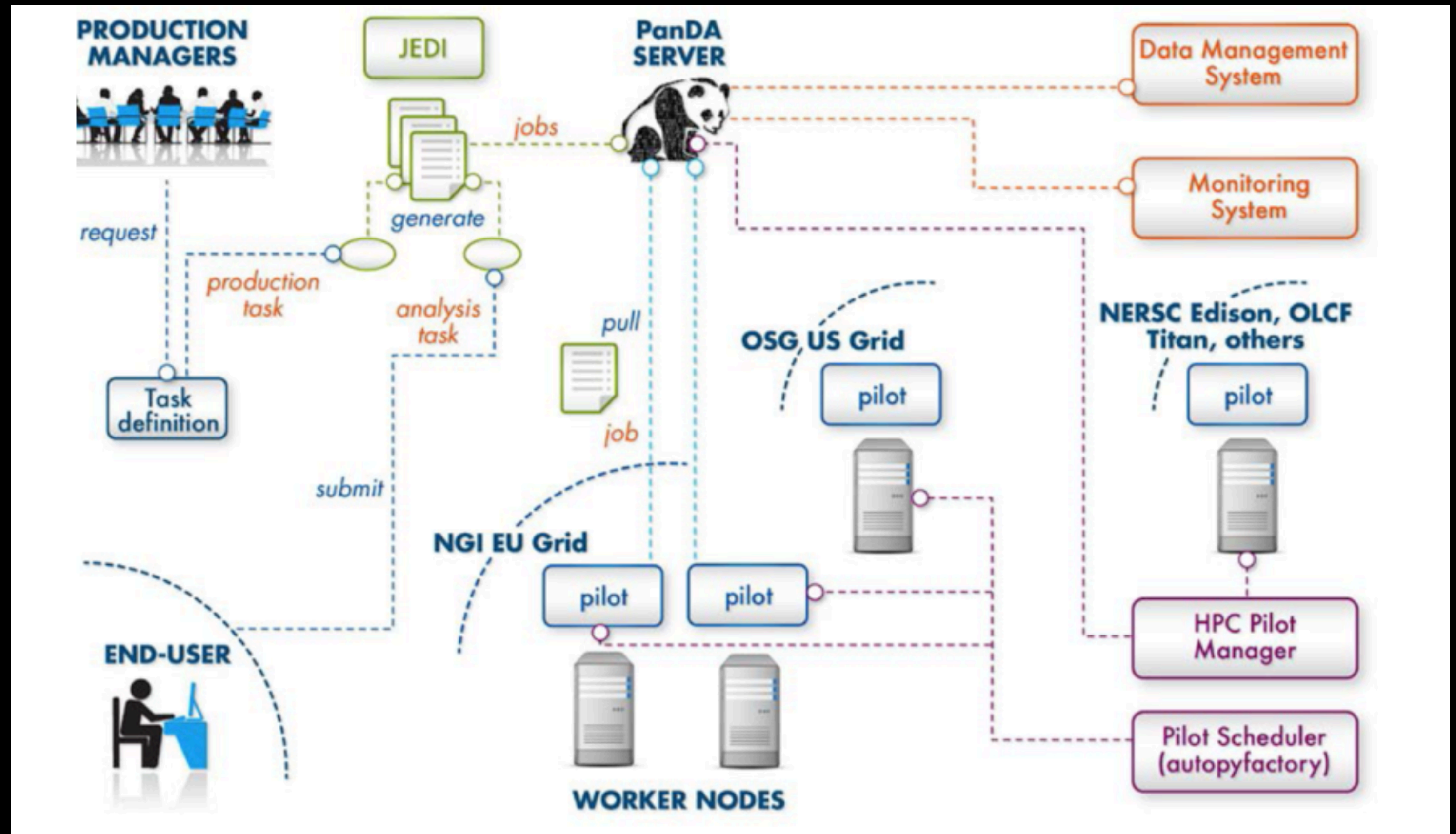
BRIEF PANDA HISTORY

- 2005: Initiated for US ATLAS (BNL and UTA)
- 2006: Support for analysis
- 2008: Adopted ATLAS-wide
- **_2009: First use beyond ATLAS_**
- 2011: Dynamic data caching based on usage and demand
- **_2012: ASCR/HEP BigPanDA project_**
- 2014: Network-aware brokerage
- 2014: Job Execution and Definition I/F (JEDI) adds complex task management and fine grained dynamic job management
- 2014: JEDI-based Event Service
- 2014: megaPanDA project supported by RF Ministry of Science and Education
- 2015: New ATLAS Production System, based on PanDA/JEDI
- 2015: Manage Heterogeneous Computing Resources
- **_2016: DOE ASCR BigPanDA@Titan project_**
- 2016: PanDA for bioinformatics
- 2016: COMPASS adopted PanDA (JINR, CERN, NRC-KI), PanDA beyond HEP : LSST, BlueBrain
- 2017: PanDA instance at OLCF: PanDA for IceCube
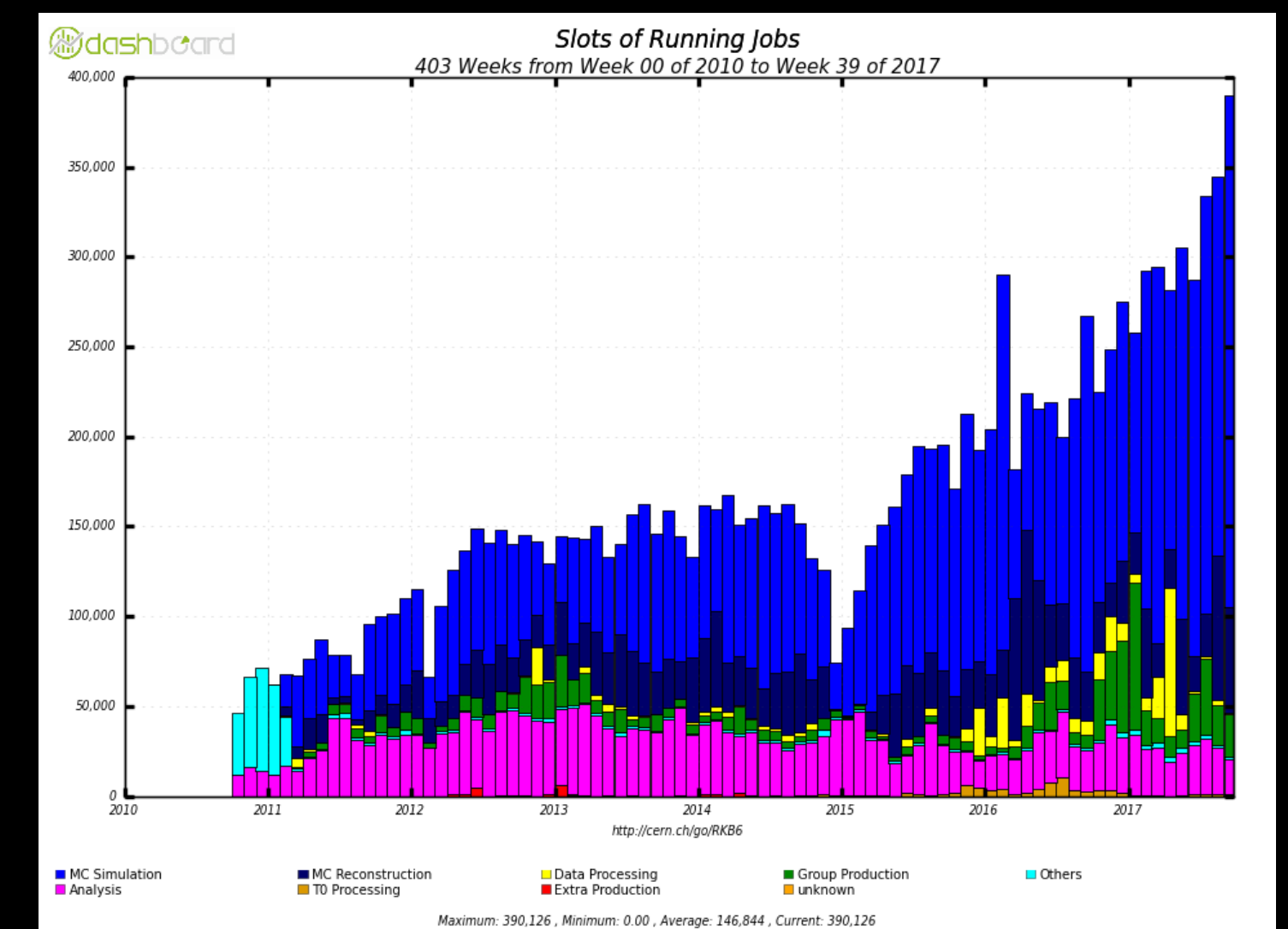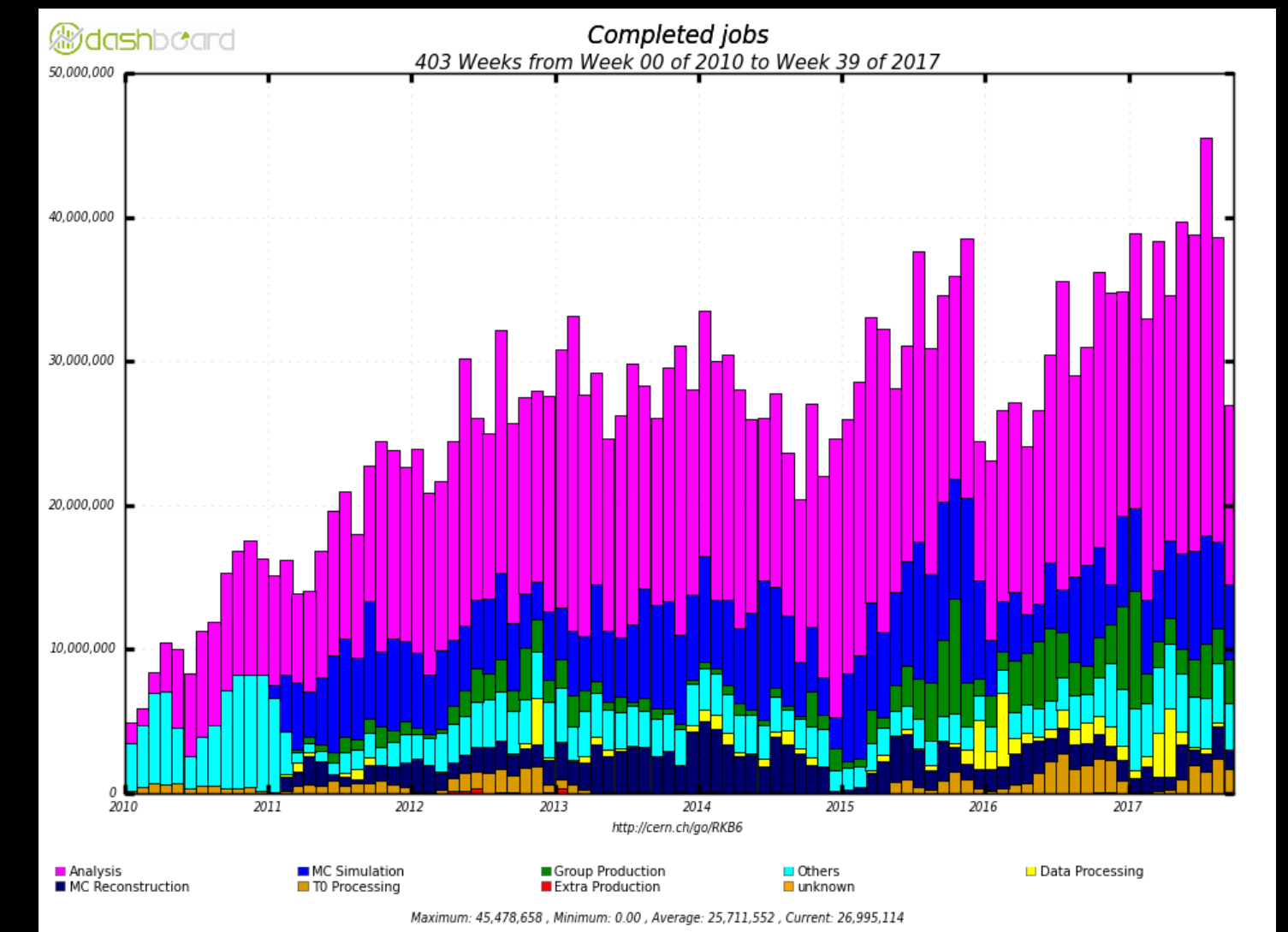
PANDA@HPC
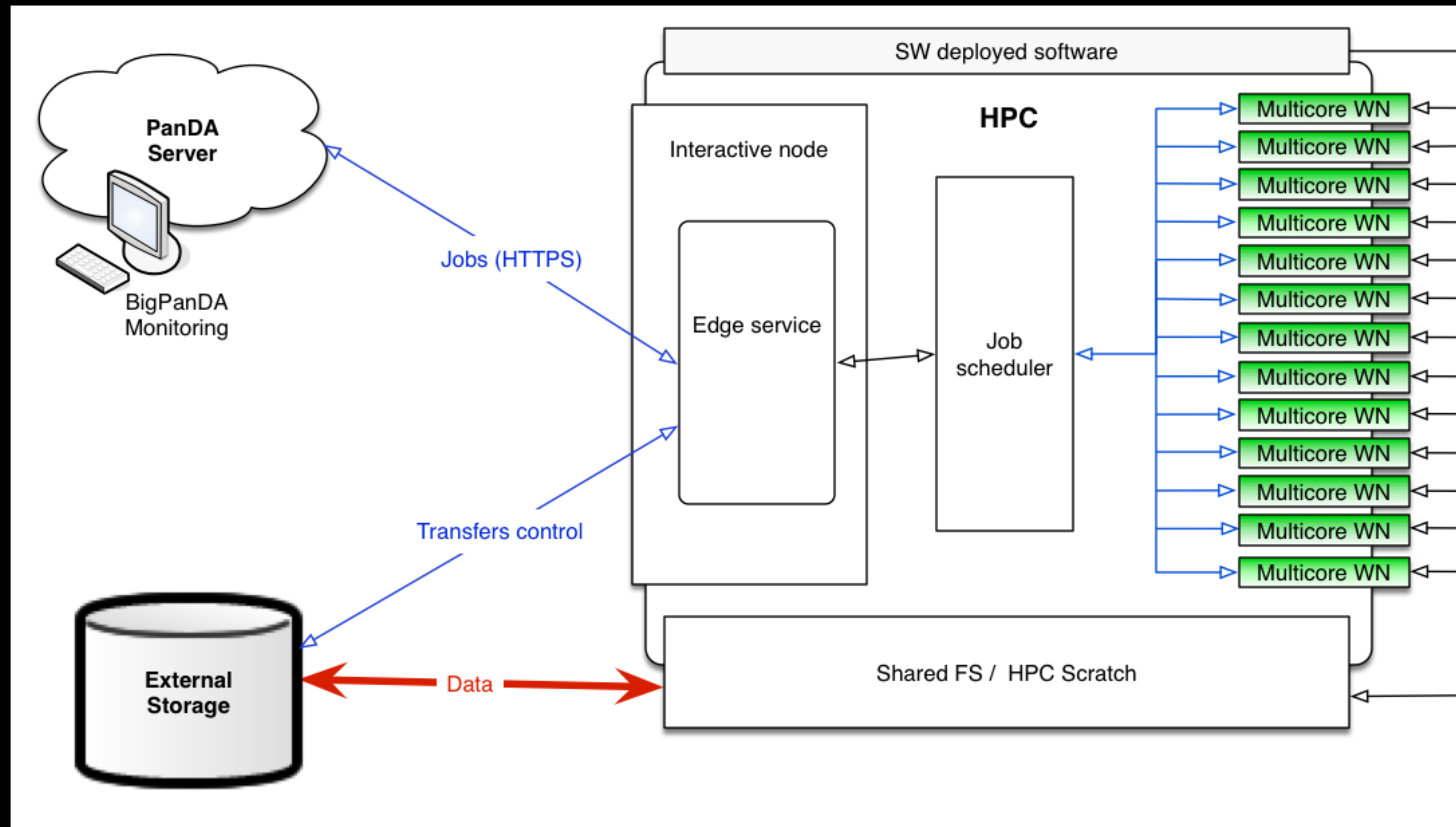
- Remote (or not so remote) PanDA server with HTTPS connection

- Edge service on interactive node - to manage interconnections with PanDA server and data motion with external storage (if needed)

- Pilot 2.0 with HPC workflow on working nodes: interconnects with edge service to manage execution of payload

MULTIVO SUPPORT IN(WITH) PANDA

- PanDA designed to support MultiVO

  - Different VO (Experiments) may share same PanDA server instance

  - Server and Pilot plugins allows to tune pre/post-processing VO specific procedures

- If VO requires high scalability (hundreds of thousands jobs per day, on wide range of resources) dedicated instance may be deployed

PANDA INSTANCE AT EC2

- During first phase of BigPanDA project PanDA instance was deployed on Amazon EC2
  - Initially used for debugging of deployment procedures and improvements for MySQL backend
  - Now used for SciDAC-4 LQCD project, LSST/DESC and nEDM job submission to supercomputers, institutional clusters and Grids
- This instance is used to support executions of workloads as on OLCF resources so on set of grid site
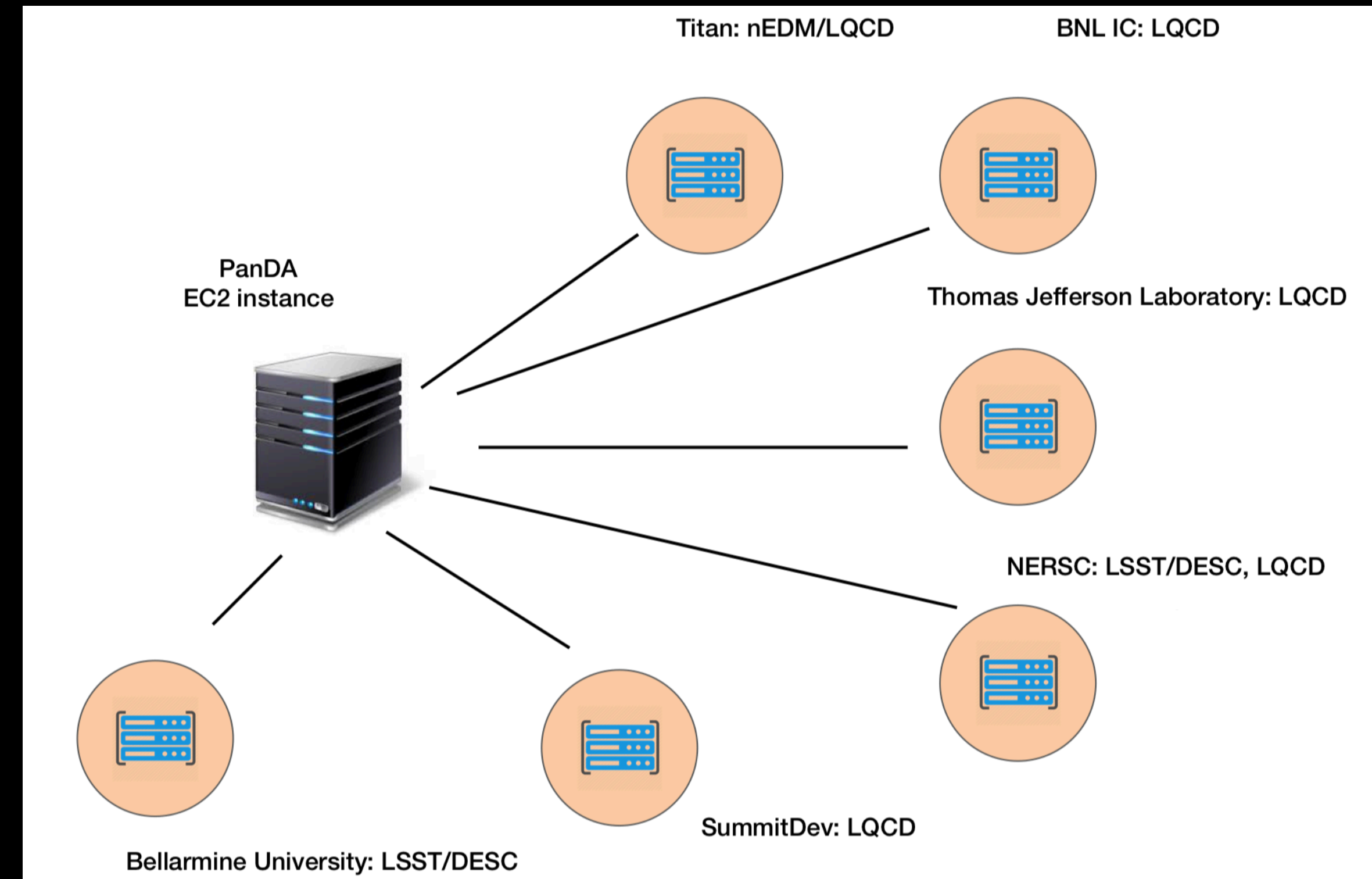
LSST - LARGE SYNOPTIC SURVEY TELESCOPE

- *The goal of the Large Synoptic Survey Telescope (LSST) project is to conduct a 10-year survey of the sky that will deliver a 200 petabyte set of images and data products that will address some of the most pressing questions about the structure and evolution of the universe and the objects in it. The LSST survey is designed to address four science areas:*
  - *Understanding the Mysterious Dark Matter and Dark Energy*
  - *Hazardous Asteroids and the Remote Solar System*
  - *The Transient Optical Sky*
  - *The Formation and Structure of the Milky Way*

- *27-ft (8.4-m) mirror, the width of a singles tennis court*
- *3200 megapixel camera (3.2 gigapixel ;-))*
- *Each image the size of 40 full moons*
- *37 billion stars and galaxies*
- *10 year survey of the sky*
- *10 million alerts, 1000 pairs of exposures,*
   *15 Terabytes of data .. every night!*
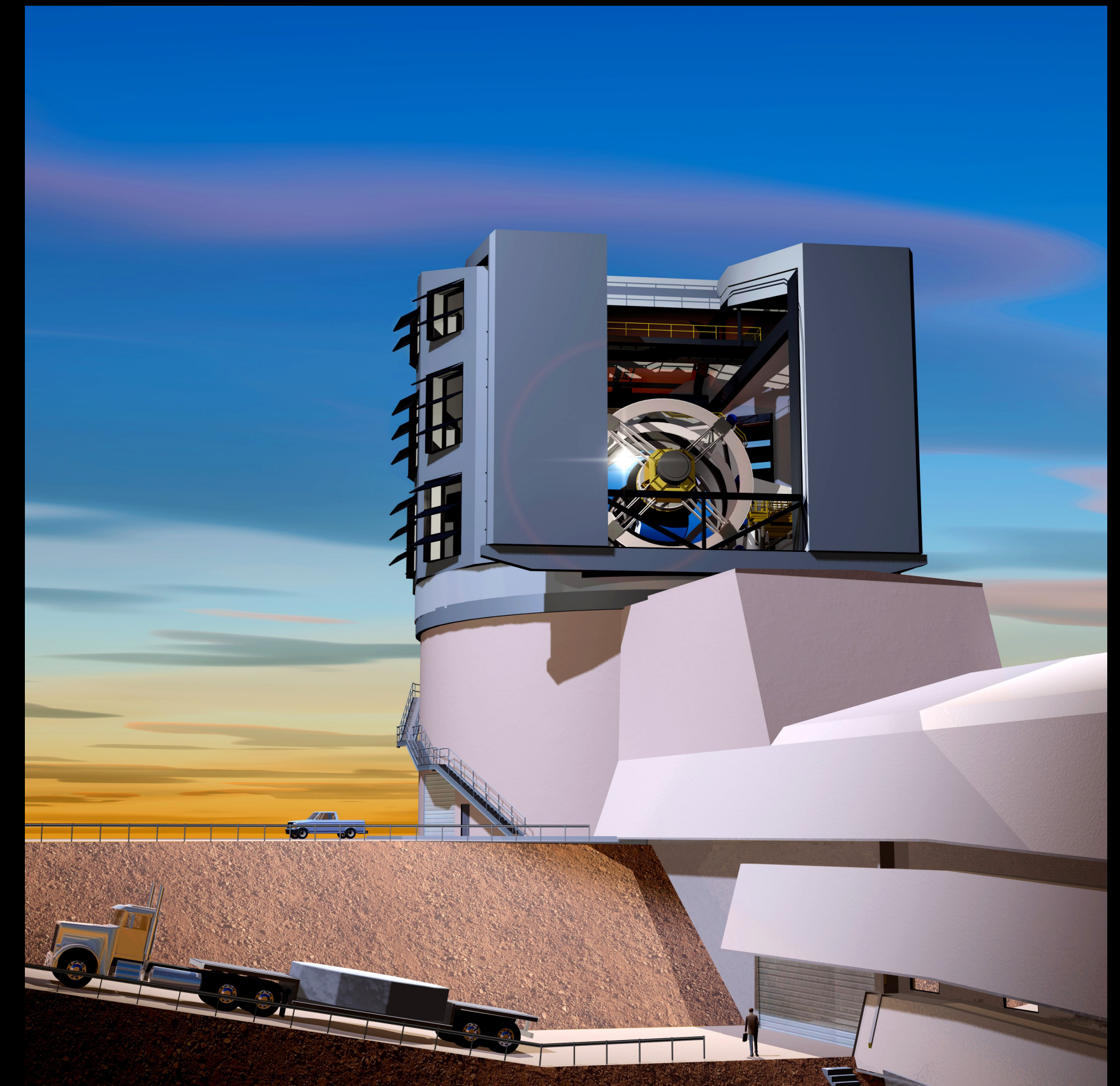
PANDA WMS FOR LSST DARK ENERGY SCIENCE COLLABORATION

- Collaboration with LSST/DESC since 2013 in terms of BigPanDA project

- The LSST Science Pipelines can process data from several telescopes using LSST's algorithms

- Pipeline to PanDA WMS submission has been implemented and tested with some 'Hello word' jobs but also with a standard DESC simulation workflow
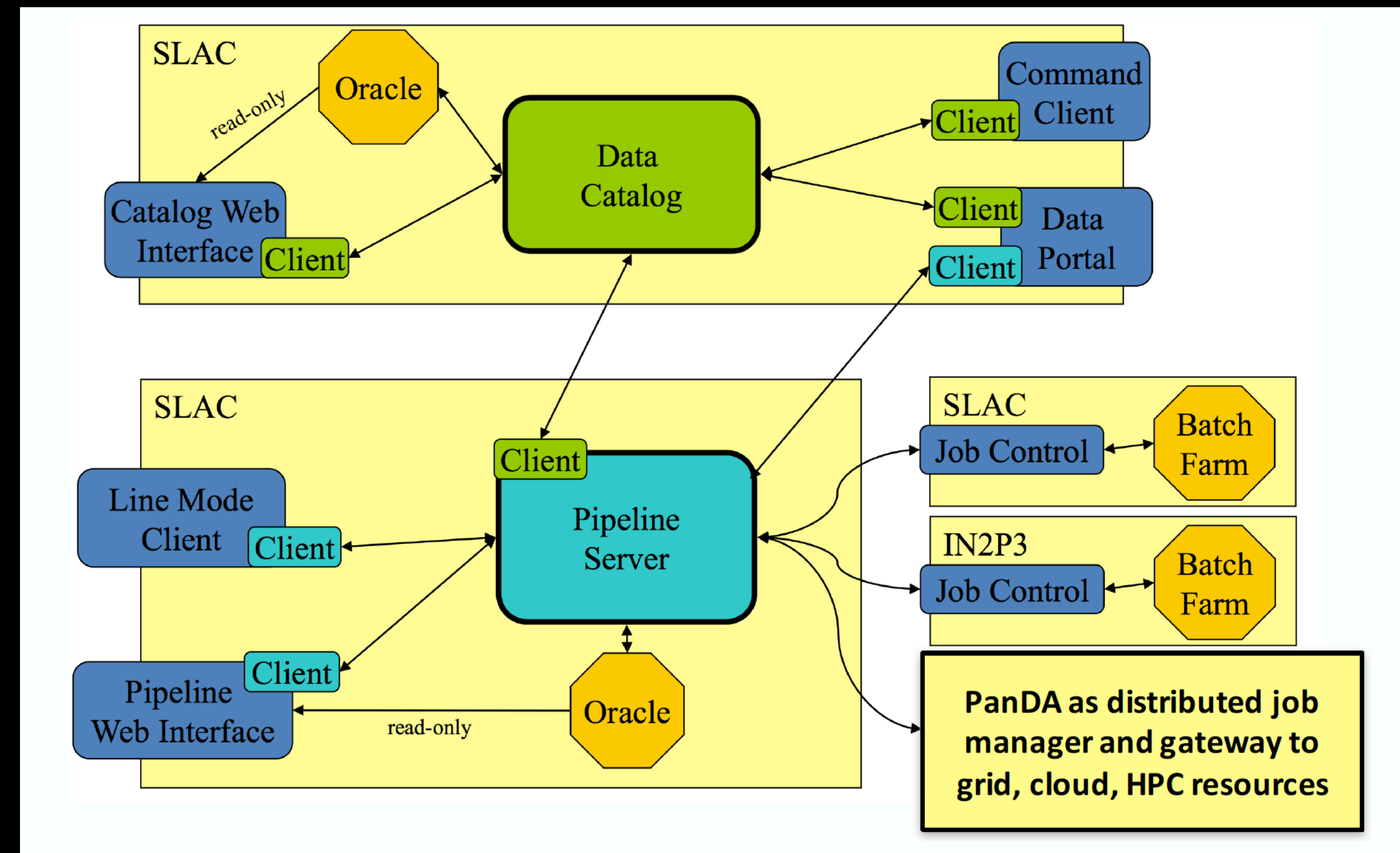
PANDA WMS FOR LSST DARK ENERGY SCIENCE COLLABORATION

- OSG: 2 sites (BNL and Bellarmine)
- GridPP
  - 31 Grid endpoints on 12 sites configured for LSST in UK,
  - 3 endpoints in France (LAPP Annecy)
- Storage for LSST now available:
  - 7 European sites (~10 TB of transient data available, data is transferred to NERSC and removed from storage)
  - 1 US (Astro storage @BNL: 200 TB)
- Ongoing process for establishing of workflow management system above PanDA WMS
- Evaluation of using of backfill resources at OLCF
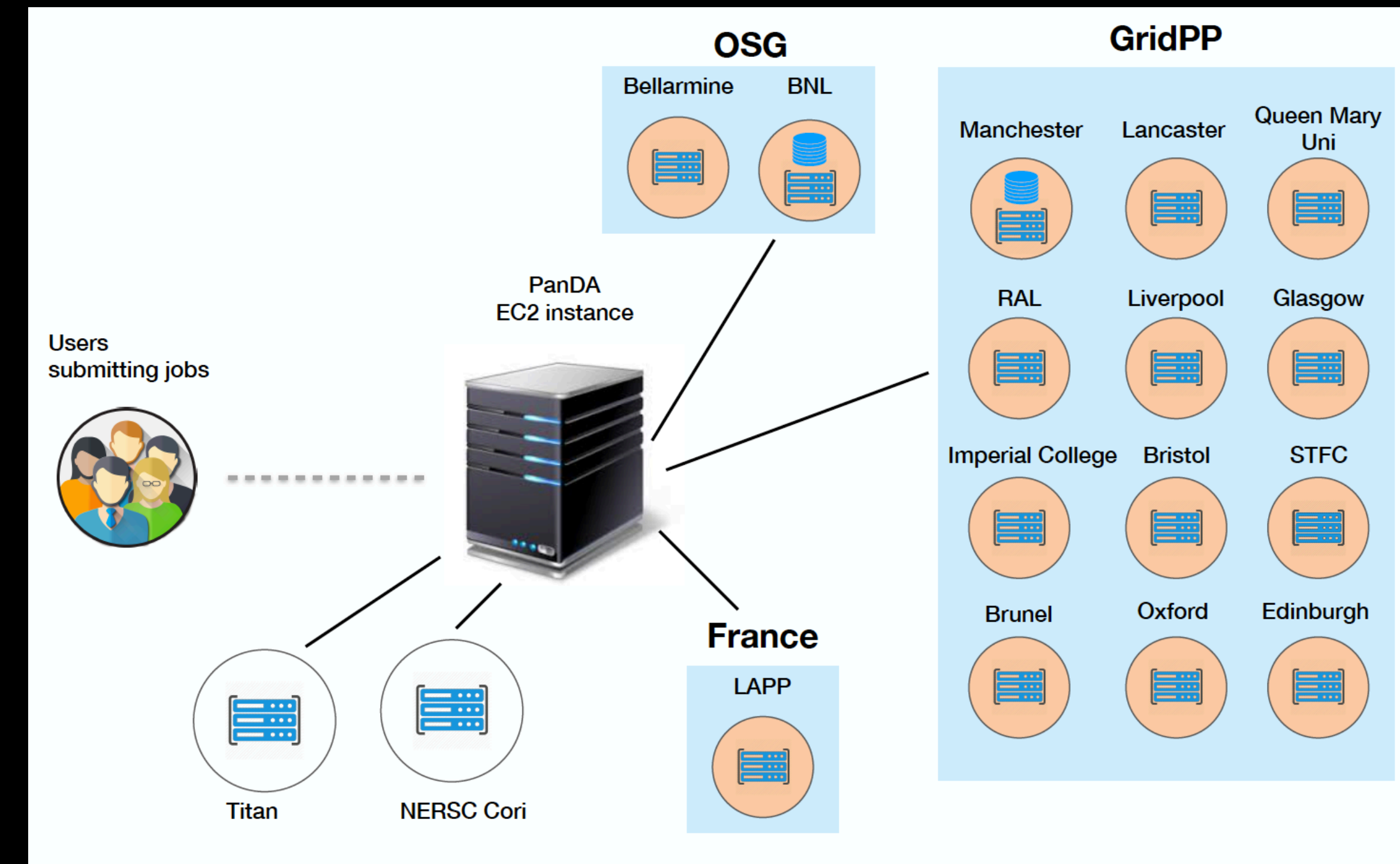
NEDM - NEUTRON ELECTRIC DIPOLE MOMENT

- 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.

- Detailed nEDM detector simulations were executed on Titan via PanDA WMS

  - Currently nEDM prepares for a future computational campaign

- nEDM PanDA is migrating from EC2 instance to OLCF instance - since only OLCF resources are going to be used for processing

  - Short time of processing of events makes nEDM good candidate for backfill consumption at OLCF
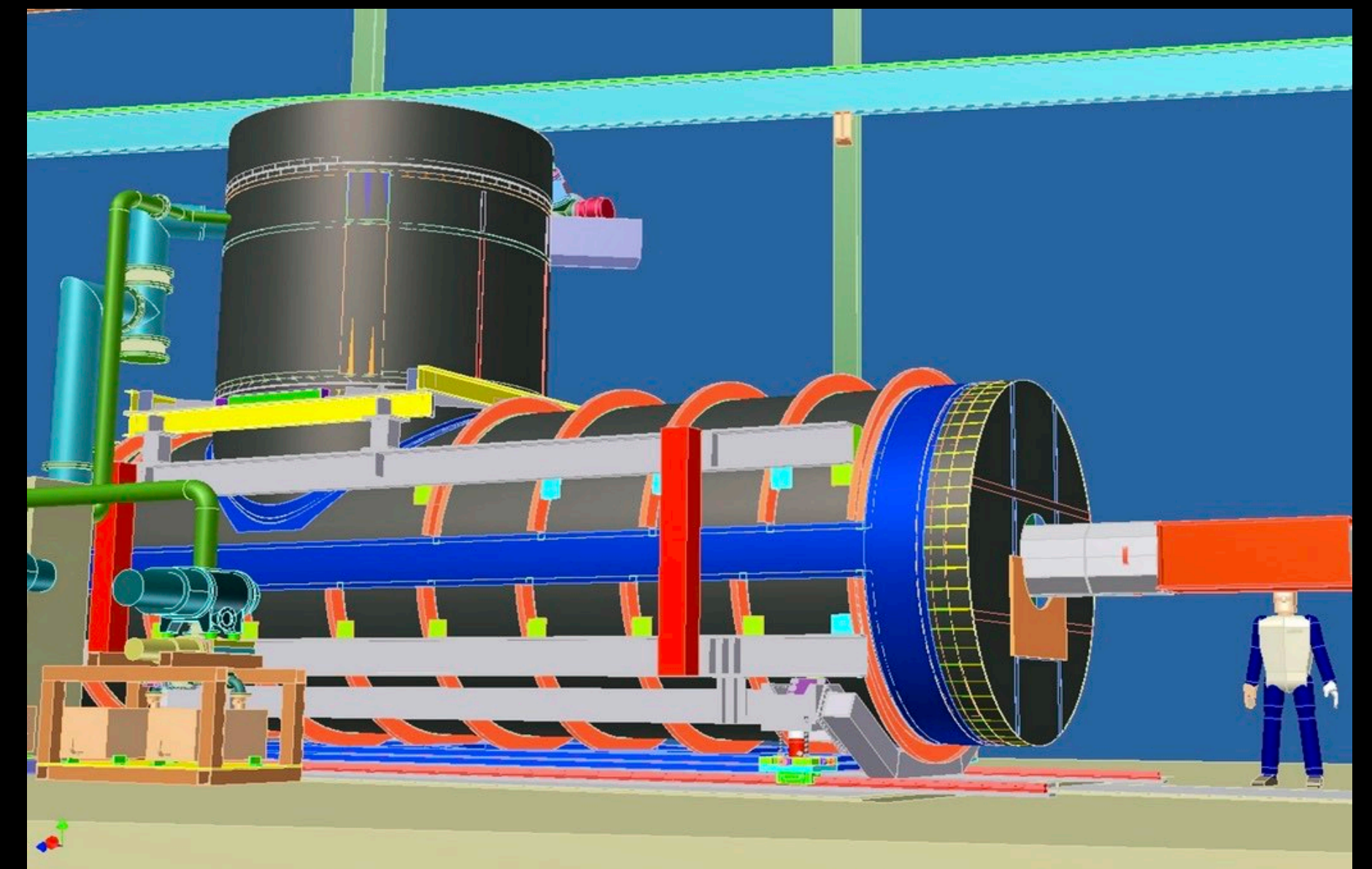
PANDA WMS FOR LATTICE QCD COMPUTATIONS

- 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. As a result, the relative costs have shifted to the point where analysis often requires an equal or greater amount of computation than gauge field generation.

- In 2017, as a part of SciDAC-4 funded project, a collaboration was formed between several US LQCD groups and BigPanDA team with the goal to adopt PanDA WMS for the needs of the SciDAC-4 LQCD computational program

PANDA WMS FOR LATTICE QCD COMPUTATIONS

- 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:

  - HPC (Titan, Cori): GPU-based, multi-node, occupying ~8000 nodes per job, ~20 hours per each job, independent jobs

  - Institutional clusters BNL, TJL: GPU-based, single-node, ~12 hours each, with workflow management

 - New kinds of payloads will be available for Summit in terms of Early Science Program
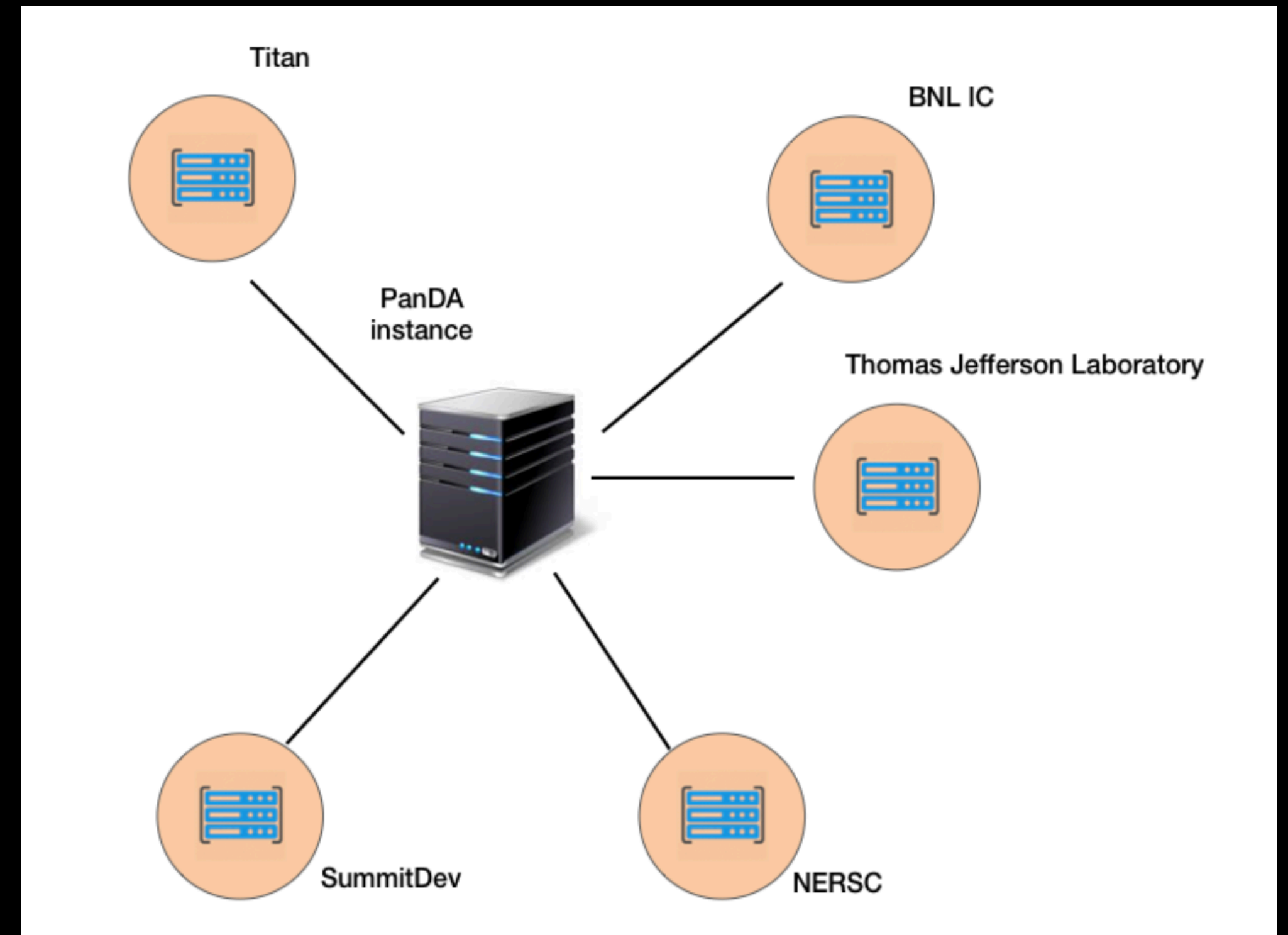
PANDA INSTANCE AT OLCF

- In March 2017, we implemented a PanDA WMS instance within ORNL operating under Red Hat OpenShift Origin - a powerful container cluster management and orchestration system in order to serve various experiments at Titan supercomputer.
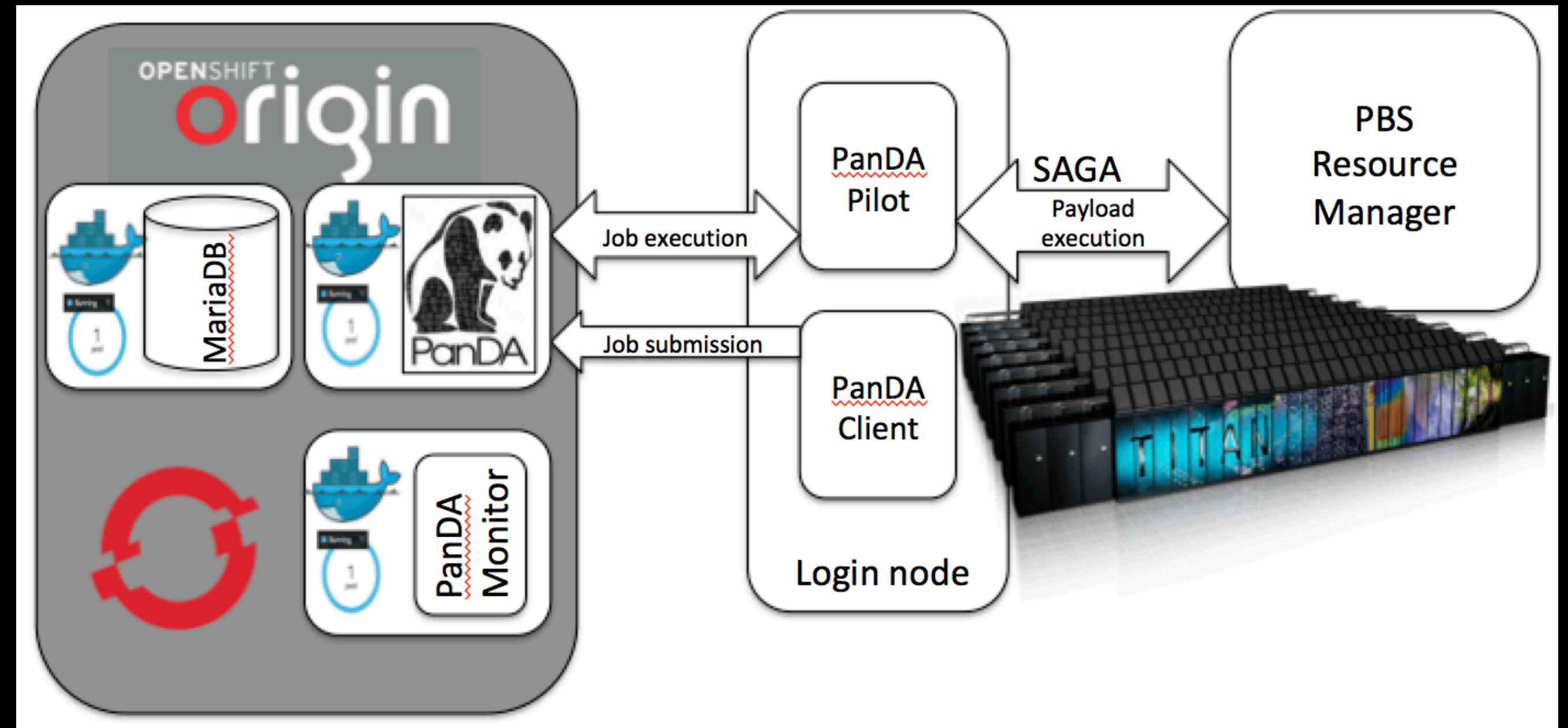
- A set of demonstrations serving diverse scientific workflows including particle physics experiments, biology studies of the genes and human brain, and molecular dynamics studies was implemented.
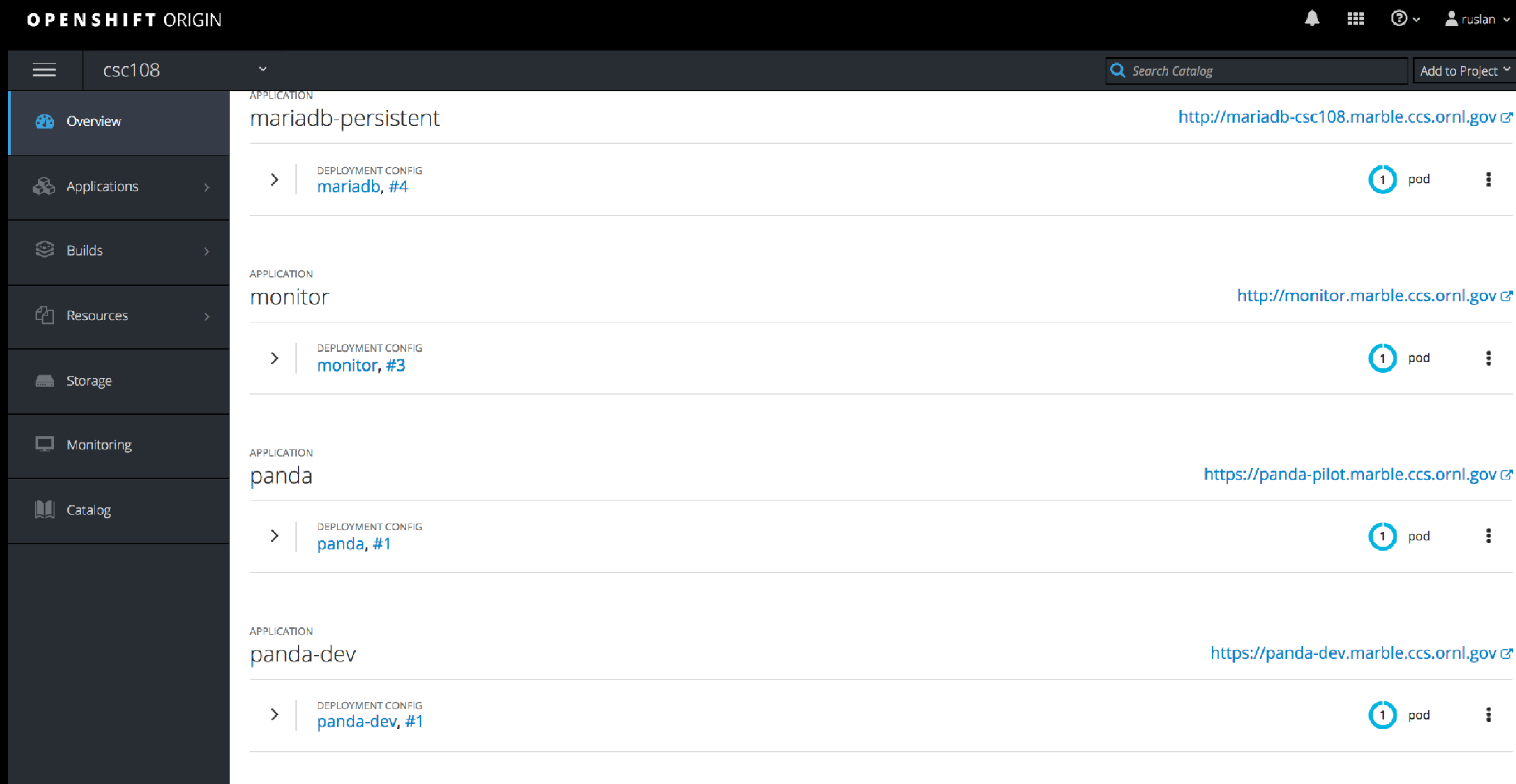
- Persistent MariaDB

- PanDA Monitor (Django)

- Production PanDA server

  - Initial implementation of edge service.

- Developer PanDA server

  - Currently working on support of the Harvester

ICECUBE NEUTRINO OBSERVATORY

- The IceCube Neutrino Observatory is the first detector of its kind, designed to observe the cosmos from deep within the South Pole ice. An international group of scientists responsible for the scientific research makes up the IceCube Collaboration

- Encompassing a cubic kilometer of ice, IceCube searches for nearly massless subatomic particles called neutrinos. These high-energy astronomical messengers provide information to probe the most violent astrophysical sources: events like exploding stars, gamma-ray bursts, and cataclysmic phenomena involving black holes and neutron stars

- IceCube collaborators address several big questions in physics, like the nature of dark matter and the properties of the neutrino itself. IceCube also observes cosmic rays that interact with the Earth's atmosphere, which have revealed fascinating structures that are not presently understood.


IceCube Collaboration
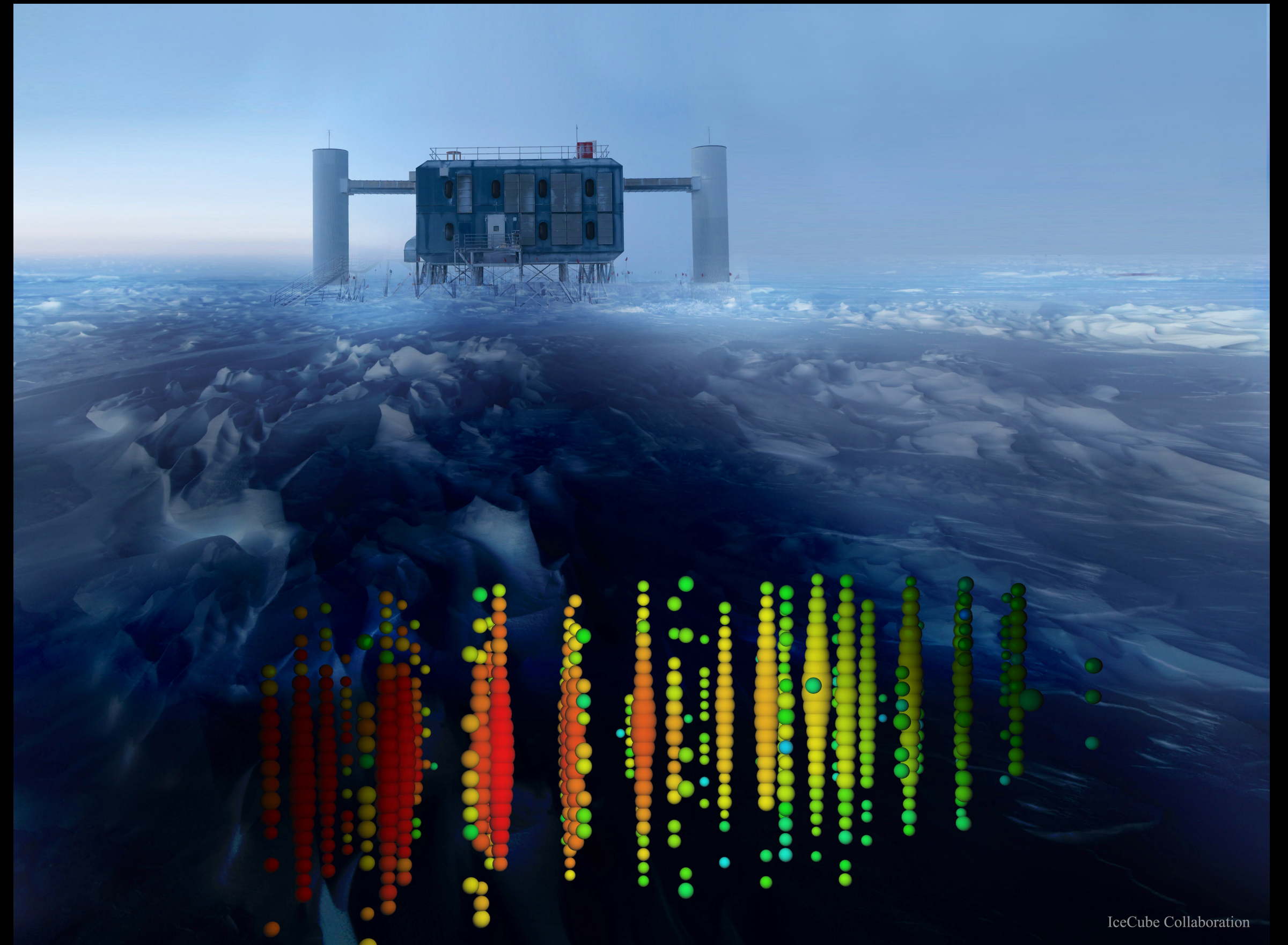
- *The IceCube detector is one cubic kilometer of ice—that would be enough water to fill one million swimming pools.*

PANDA WMS FOR ICECUBE

- Together with experts from the experiment we're working on the demonstration of a real IceCube workflow Titan

  - Application: NuGen package - GPU application for atmospheric neutrinos simulations and analysis

  - Application packed in Singularity container.

  - Whole node, but not MPI application:

- ~45000 jobs in campaign (5000 input files)

  - Remote stage-in/-out the data from GridFTP storage with GSI authentication

ICECUBE PAYLOADS ON TITAN

- IceCube payloads were not designed initially for support of MPP, so execution of this kind of payload in the traditional way of HPC will be inefficient (one node per job is not good for LCF's policies)

- PanDA on HPC allows to combine this kind of payload into assemblies and executes them as a simple MPI application (it was required to get efficient execution of ATLAS payloads)

  - To be effective this approach is required to join payloads with similar wall time into assemblies

  - Unfortunately, IceCube payloads do not have this type of characterization
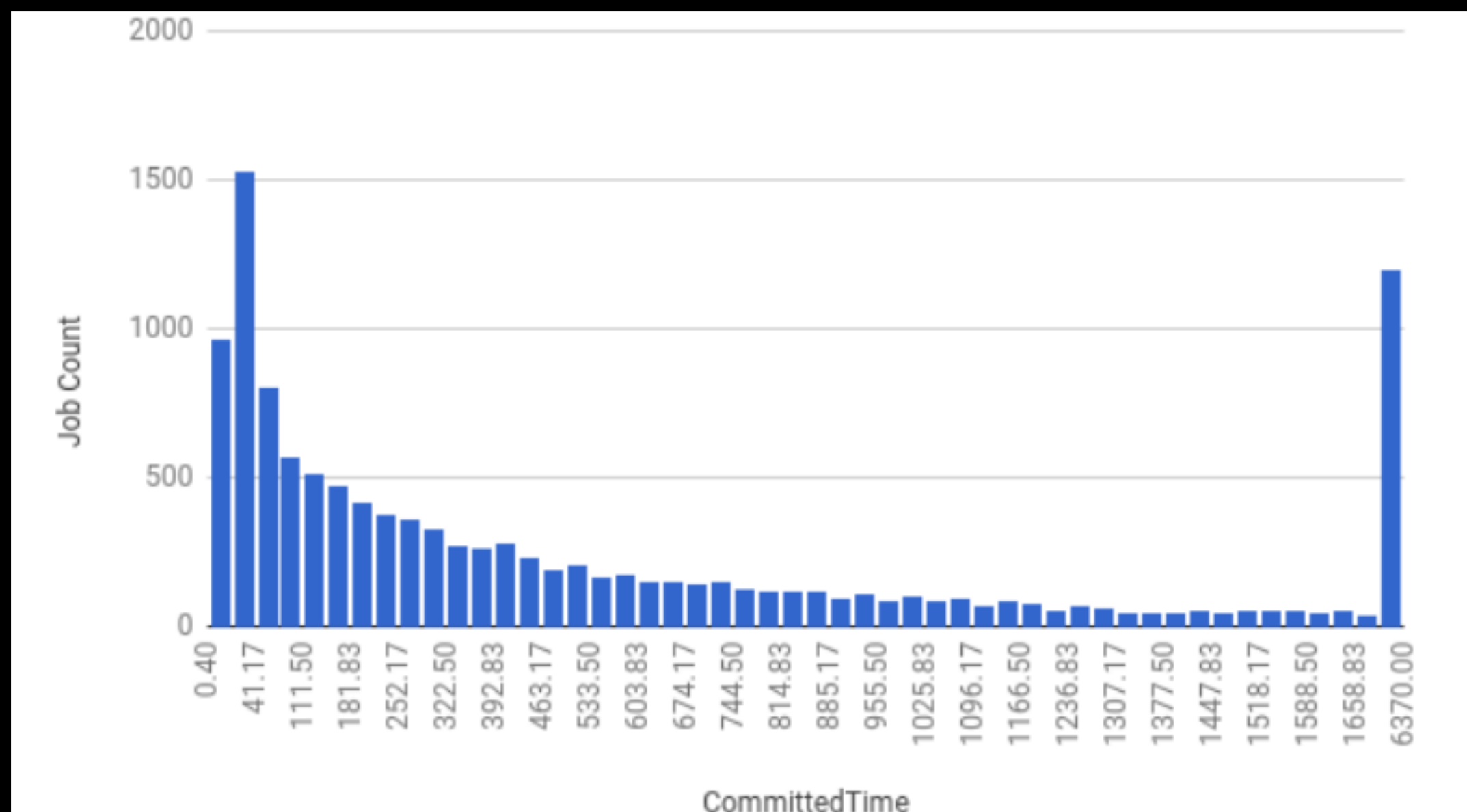

- To optimize quota usage, the processing will be conducted in a few steps.

- On each step, we will run multiple job processes in parallel via MPI wrapper

- Starting with the walltime=20min on the first step, all failed jobs (jobs not completed because of the walltime limit) will be resubmitted on the next step with longer walltime

- Simulating Enzyme Catalysis, Conformational Change, and Ligand Binding/Release. (Prof. Kwangho Nam (University of Texas at Arlington, USA)


Multiscale simulation

Protein kinases
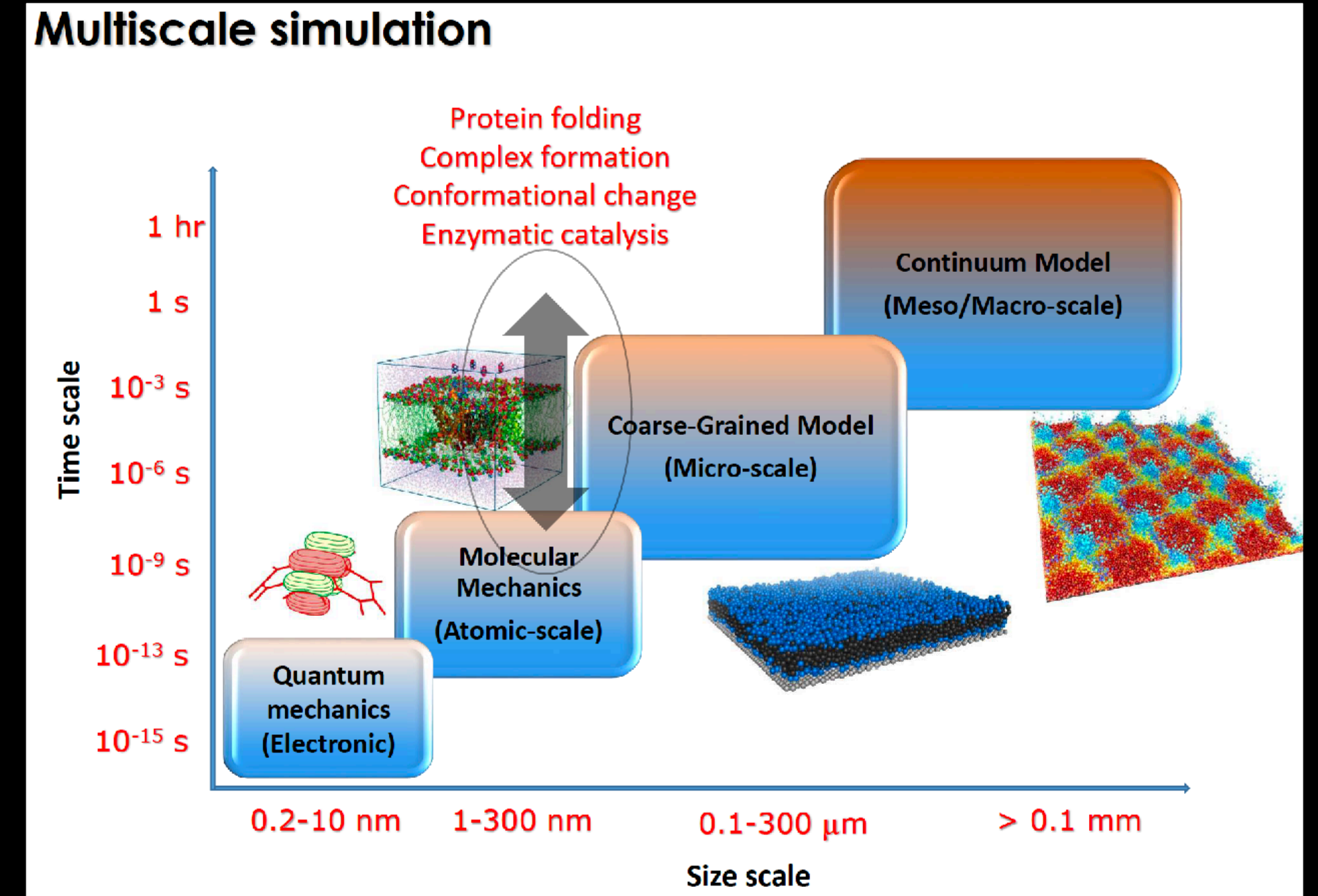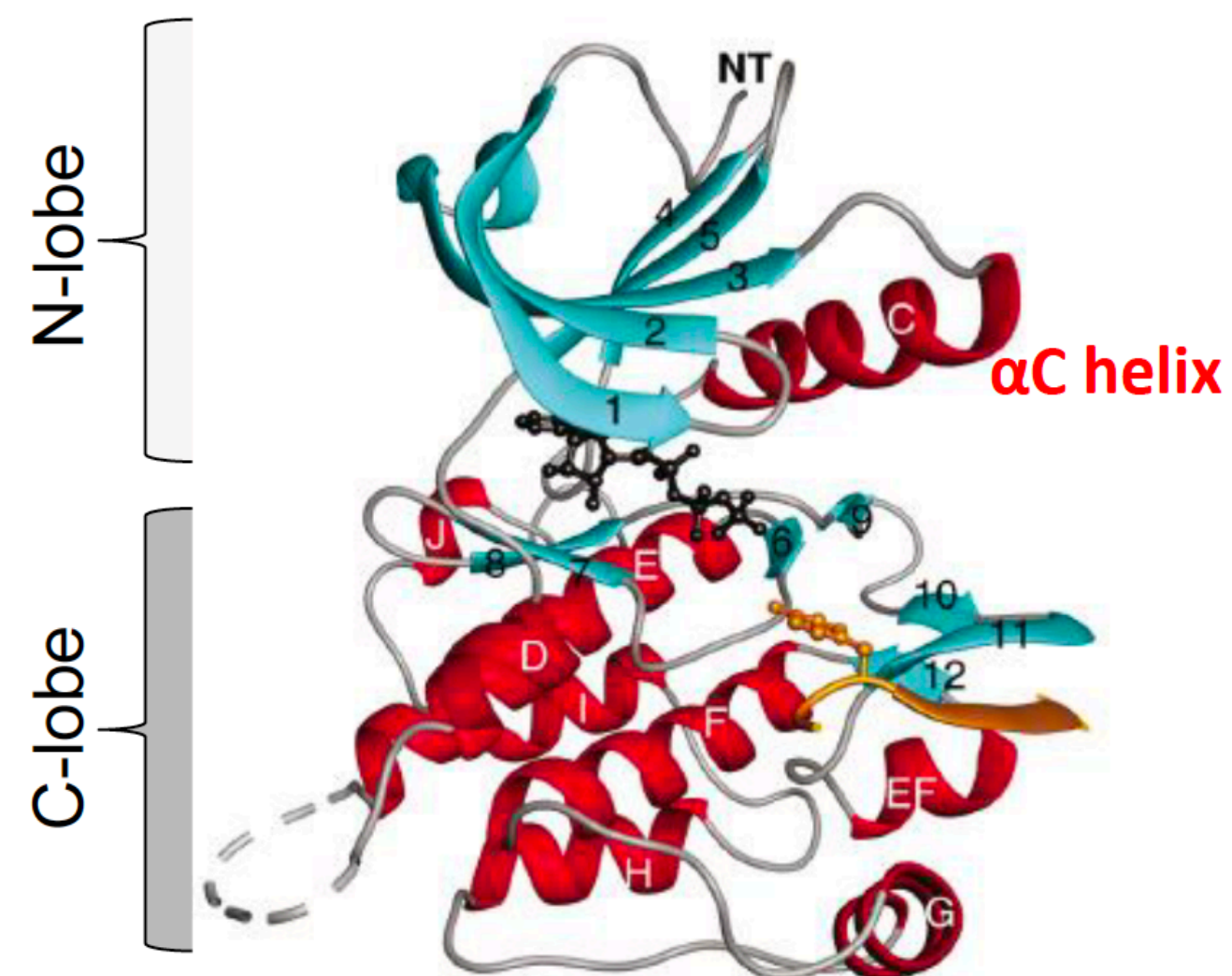
- CHARMM (Chemistry at HARvard Macromolecular Mechanics) payload (hybrid MPI/OpenMP/GPU) example built and executed on Titan

- Depending on the type of projects, payloads can expand beyond 500 nodes on Titan; currently, it uses 60-124 nodes for each project

- 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', i.e. the genetic background.

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

- Input data were located in a set of eight input directories of 152 M each. Every PanDA job was configured to process single input directory in backfill mode on one node and walltime of 30 min. The output data of about 11M per job was located to the corresponding output dir

Nature Reviews | Genetics
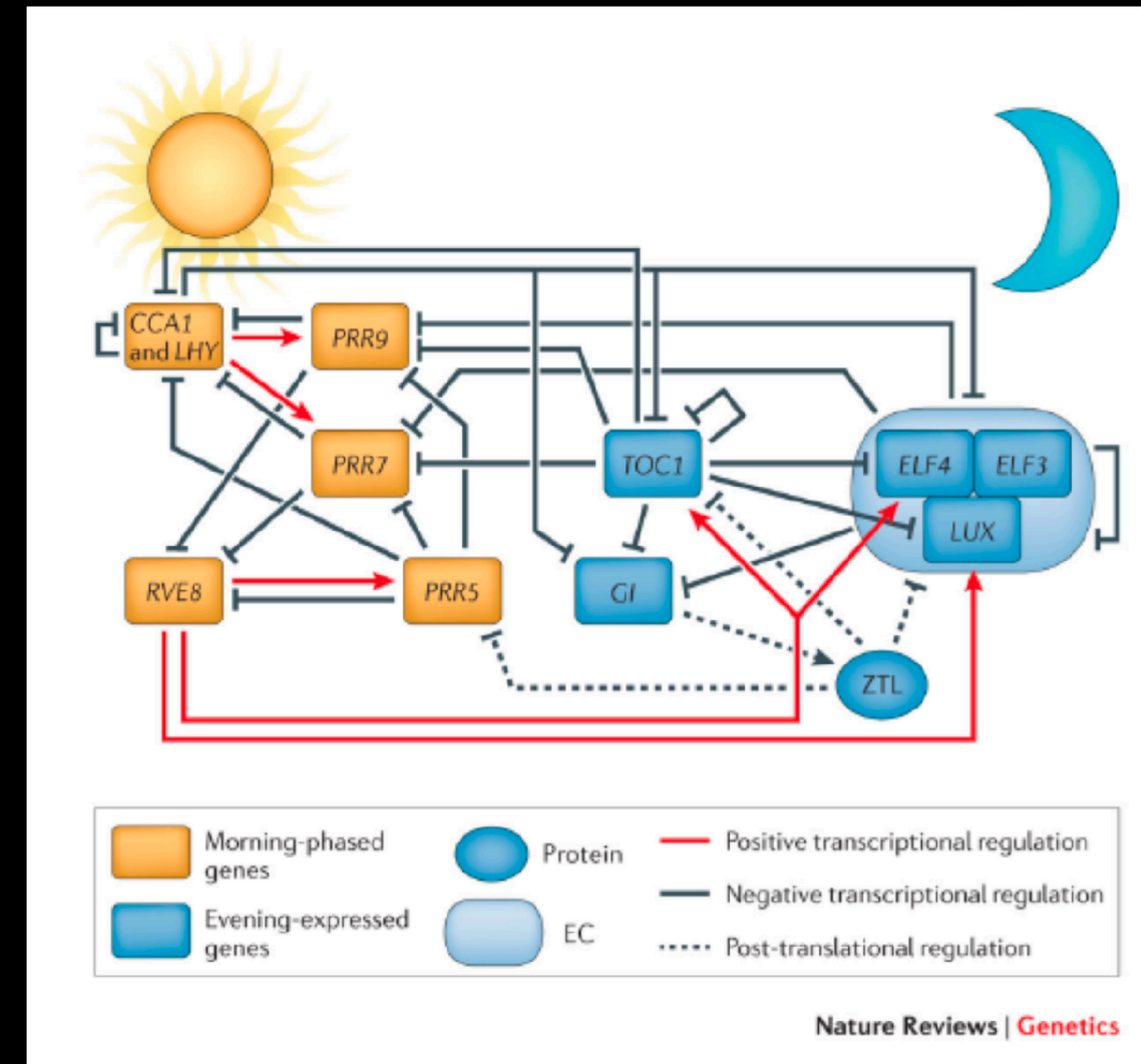
- Genetic regulatory networks
  - Phenotypes, or traits, are determined by genetic regulatory networks
  - These genetic networks are composed of genes that are organized to coordinate overt phenotypes
  - In quantitative genetic analyses, genes underlying the basis of traits are formalized

THE CENTER FOR BIOENERGY INNOVATION

BLUE BRAIN PROJECT

- The goal of the Blue Brain Project is to build biologically detailed digital reconstructions and simulations of the rodent, and ultimately the human, brain

- The supercomputer-based reconstructions and simulations built by the project offer a radically new approach for understanding the multilevel structure and function of the brain

- The novel research strategy allows the project to build computer models of the brain at an unprecedented level of biological detail

- Supercomputer-based simulation turns understanding the brain into a tractable problem, providing a new tool to study the complex interactions within different levels of brain organization and to investigate the cross-level links leading from genes to cognition
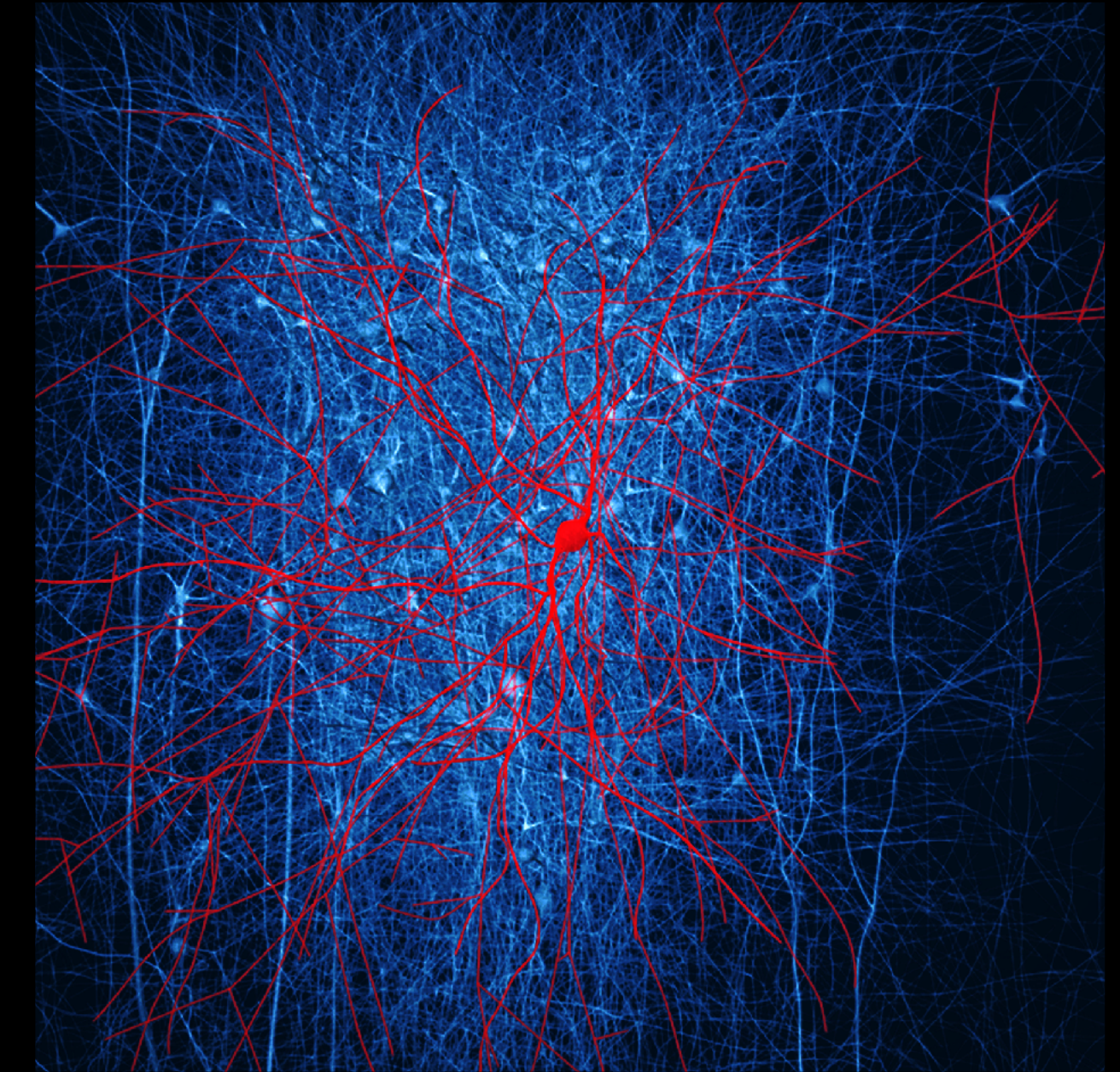
THE PRESYNAPTIC NEURONS OF A LAYER 2/3 NEST BASKET CELL (IN RED) WERE STAINED (IN BLUE) IN A DIGITAL RECONSTRUCTION OF NEOCORTICAL MICROCIRCUITRY. ONLY IMMEDIATE NEIGHBOURING PRESYNAPTIC NEURONS ARE SHOWN.

BLUE BRAIN PROJECT COMPUTING MODEL EVOLUTION WITH PANDA WMS
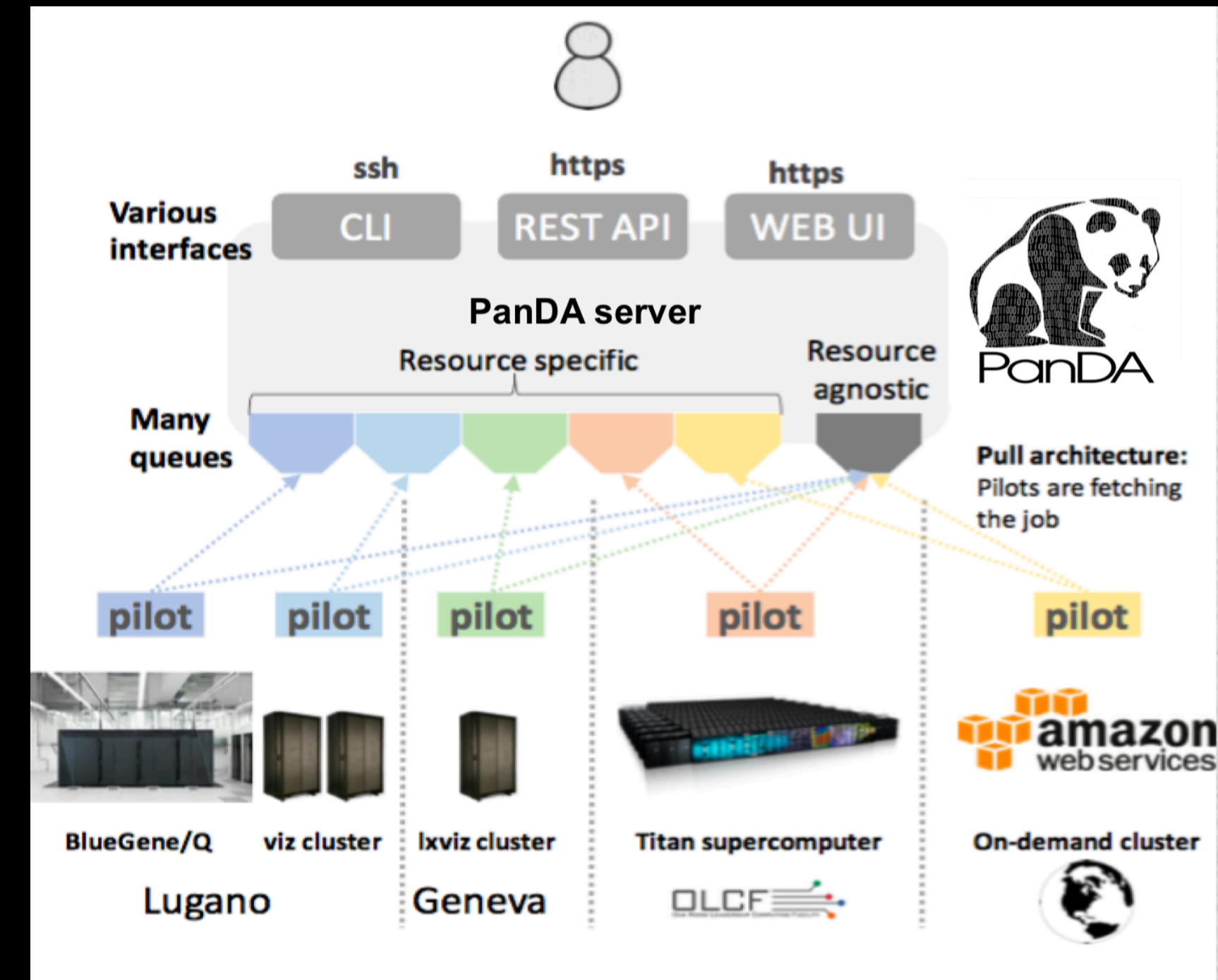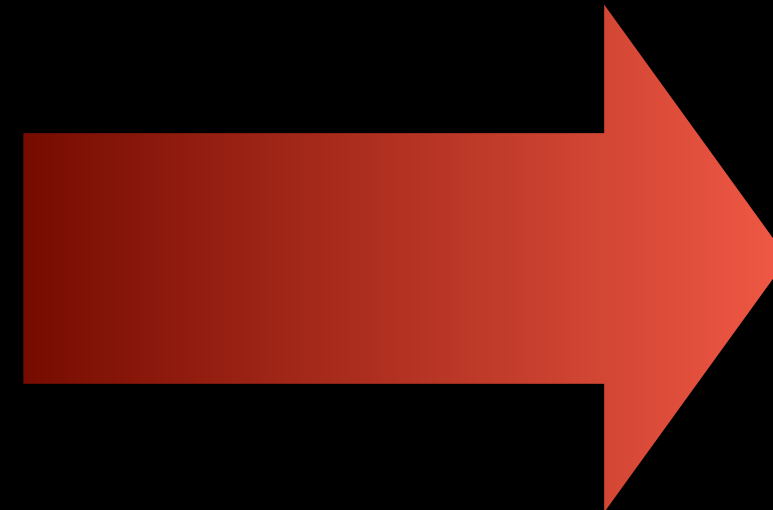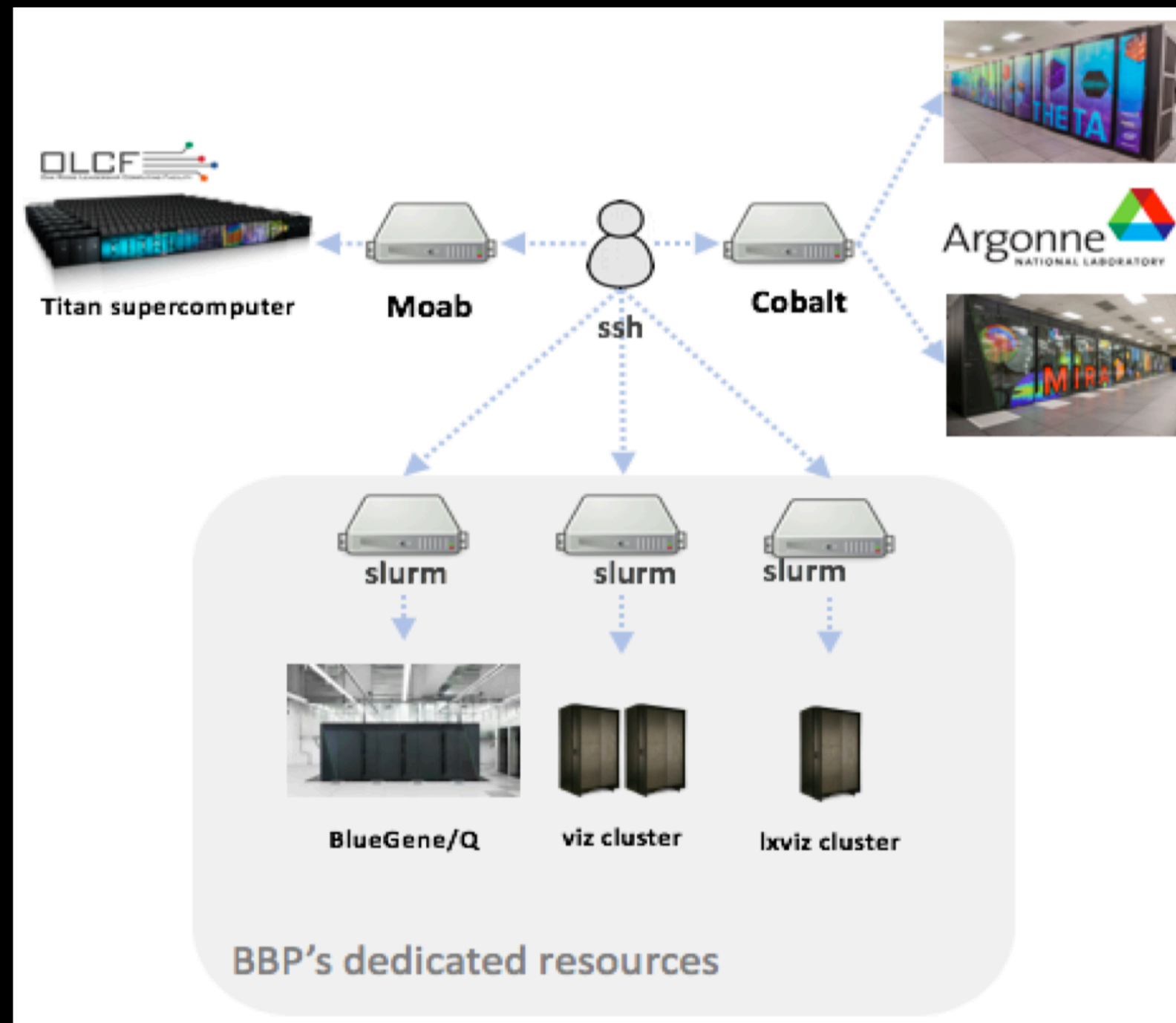
- Dedicated PanDA instance on VM hosted by BBP
  - Improved PanDA User Interface: from CLI to Web interface
  - Integration with BB authentication system

PANDA WMS FOR BLUE BRAIN PROJECT

- Phase 1 of "proof of the concept" was successfully finished
  - The test jobs were successfully submitted to the targeted resources via PanDA portal.
  - The project demonstrated that the software tools and methods for processing large volumes of experimental data, which have been developed initially for experiments at the LHC, can be successfully applied to BBP.
- Phase 2 of "pre-production" is under investigation currently

SUMMARY

- BigPanDA project demonstrates:

  - possibility to improve integration with heterogeneous computing resources

  - extend automation of processing of significant amount of data

  - possibility of integration with workflow management systems for orchestration of complex data processing