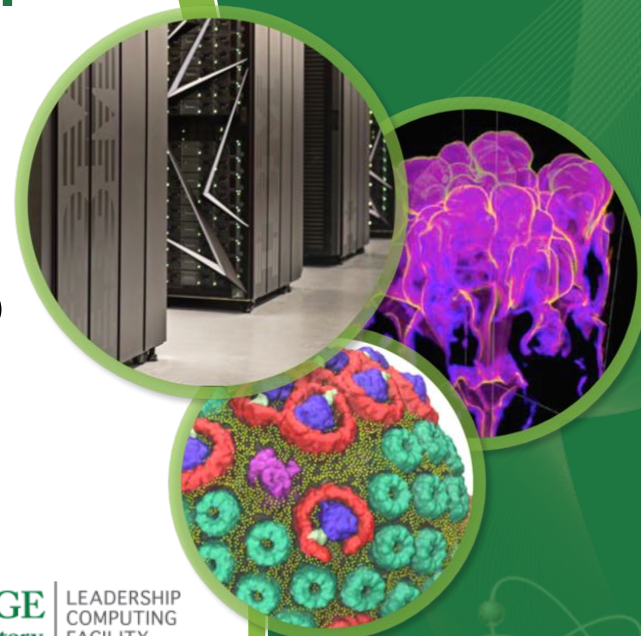
BigPanDA Workflow Management on Titan for High Energy and Nuclear Physics and for Future Extreme Scale Scientific Applications

Alexei Klimentov (PI, BNL), Kaushik De (Co-PI, UT-Arlington),
Jack Wells (Co-PI, ORNL), Shantenu Jha (Co-PI, Rutgers/BNL)



Next-Generation Networks for Science PI Meeting

Gaithersburg

25 September 2018

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

BROOKHAVEN NATIONAL LABORATORY

UNIVERSITY OF TEXAS ARLINGTON

THE STATE UNIVERSITY OF NEW JERSEY RUTGERS

ORNL is managed by UT-Battelle
for the US Department of Energy

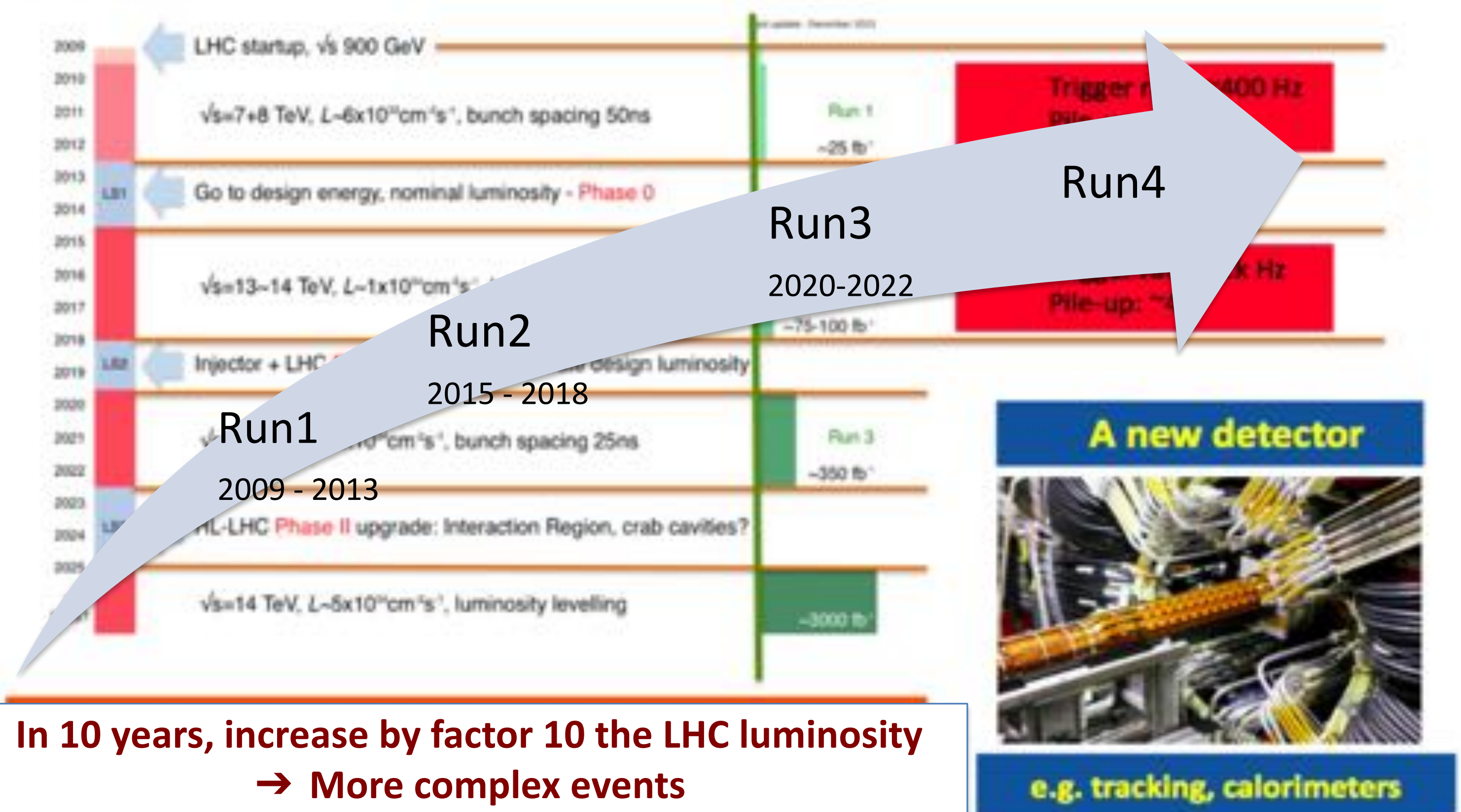
This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Outline

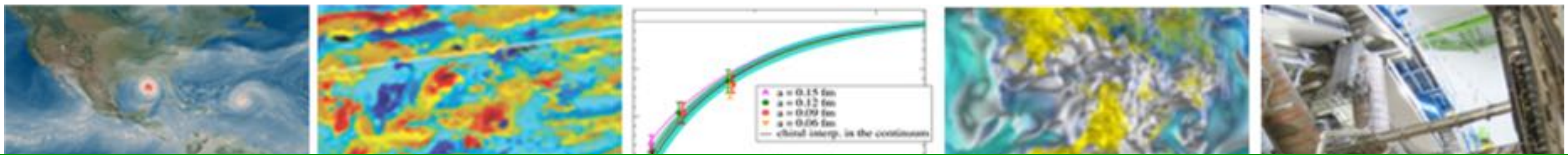
- Science requirements for Experimental and Observational Data (EOD) Science: highlight DOE/SC/ASCR workshop reports.
- LHC/ATLAS – OLCF integration: Big PanDA Demonstrator project at OLCF
- Impact of BigPanDA project on HEP/ATLAS Computing
- PanDA WMS beyond HEP: PanDA server instance at OLCF
- Impact of BigPanDA project on OLCF support of user-managed middleware services.
- Conclusions

Experimental and Observational Science Data is Exploding

LHC Upgrade Timeline



In 10 years, increase by factor 10 the LHC luminosity
➔ More complex events
➔ More Computing Capacity



DOE-ASCR Exascale Requirements Reviews

ASCR facilities conducted six exascale requirements reviews in partnership with DOE Science Programs

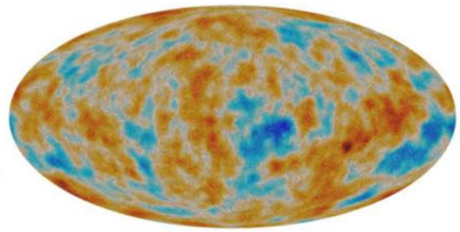
- Goals included:
  - Identify mission science objectives that require advanced scientific computing, storage and networking in exascale timeframe
  - Determine future requirements for a computing ecosystem including data, software, libraries/tools, etc.

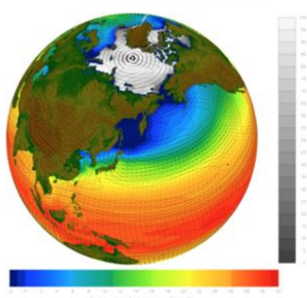
Schedule

June 10–12, 2015	HEP
November 3–5, 2015	BES
January 27–29, 2016	FES
March 29–31, 2016	BER
June 15–17, 2016	NP
Sept 27–29, 2016	ASCR
March 9–10, 2017	XCut

All 7 workshop reports are available online: http://exascaleage.org/

Common Themes Across DOE Science Offices
Data: Large-scale data storage and analysis

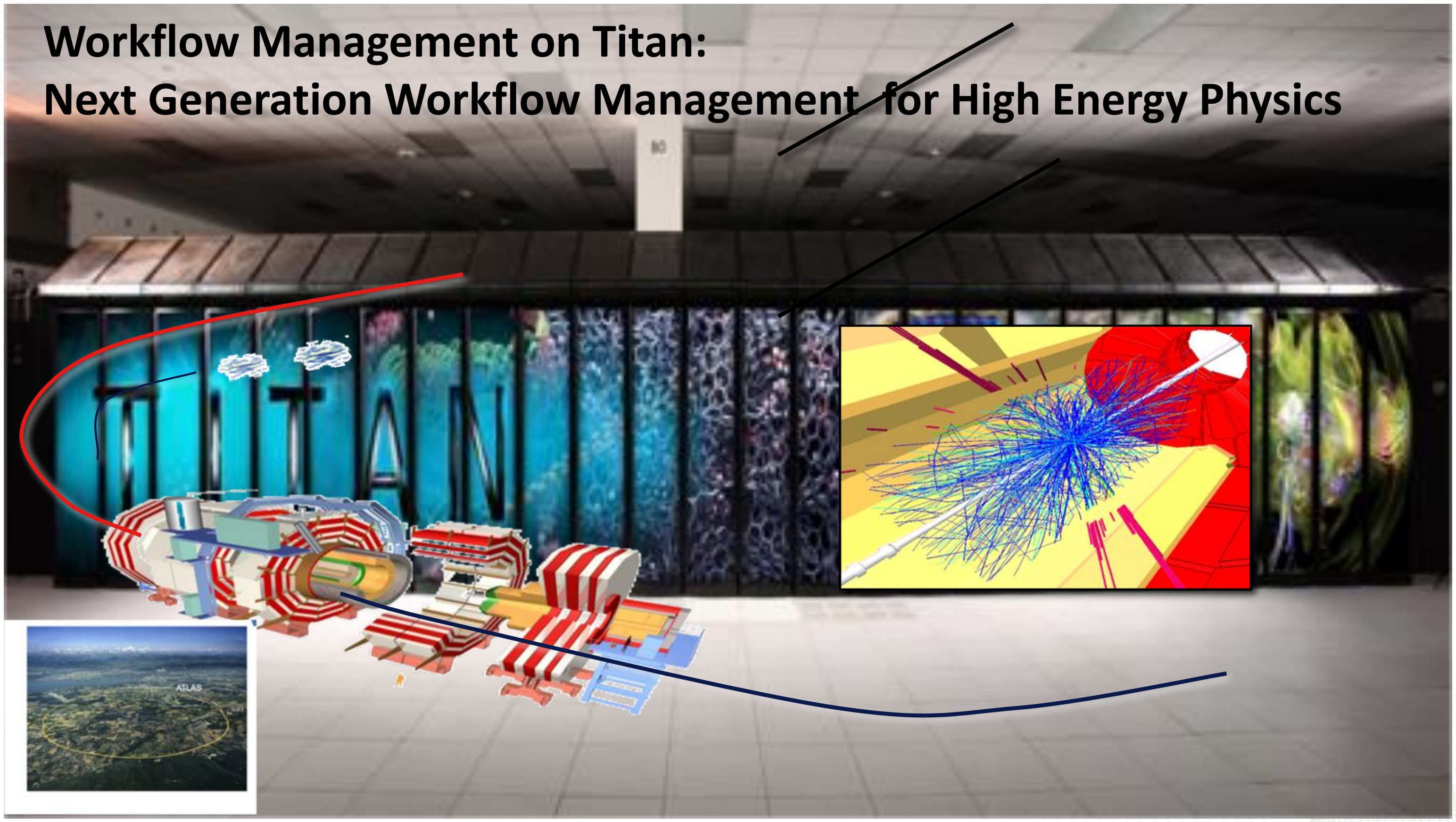




Experimental and simulated data set volumes are growing exponentially.
Examples: <u>High luminosity LHC</u>, light sources, climate, cosmology data sets ~ 100s of PBs.
Current capability is lacking.

<u>Methods and workflows of data analytics are different than those in traditional HPC.</u>
Machine learning is revolutionizing field.
Established analysis programs must be accommodated.

LHC/ATLAS-OLCF Integration



"BigPanDA Workflow Management on Titan for High Energy and Nuclear Physics and for Future Extreme Scale Scientific Applications," DOE/SC/ASCR Next-Generation Networking for Science, Rich Carlson.
PI: Alexei Klimentov (BNL); Co-Pis; K. De (U. Texas-Arlington), S. Jha (Rutgers U) J.C. Wells (ORNL)

Research Challenges

- Motivation:
    - Demand for resources greatly outstrips supply,
    - New payloads are being developed well suited for HPC architectures
- Challenge: Resource management approaches, policies and software are different for HPC and Grids
- Objectives:
    - How can we provide existing workload management capabilities (PanDA) uniformly across HPC and Grids?
    - Use it to utilize otherwise unusable resources on HPC as well as traditional allocations queue-based resource management

The Opportunity for HPC-Grid (HTC) Integration I

How do we efficiently integrate HPC resources and distributed High-Throughput Computing (HTC, or Grid) resources?

- From the perspective of large supercomputer centers, how best to integrate large capability workloads, e.g., the traditional workloads of leadership computing facilities, with the large capacity workloads emerging from, e.g., experimental and observational data?
- Workload Management Systems (WLMS) are needed to effectively integrate computing for experimental and observation data into our data centers.

The Opportunity for HPC-Grid (HTC) Integration II

The ATLAS experiment provides an attractive science driver, and the PanDA Workflow Management System has attractive features for capacity-capability integration

- *The Worldwide LHC Computing Grid and a leadership computing facility (LCF) are of comparable compute capacity.*
  - *WLCG: 700,000's x86 compute cores*
  - *Titan: 300,000 x86 compute cores and 18,000 GPUs*
- *There is a well-defined opportunity to increase LCF utilization by increasing backfill opportunities.*
  - *Batch scheduling prioritizing leadership-scale jobs results in ~90% utilization of available resources at DOE's LCFs.*
  - *Up to 10% of Titan's cycles (~400M core hours) are available if a very large volume of capacity jobs can be run in backfill mode.*

Deploying PanDA on Titan

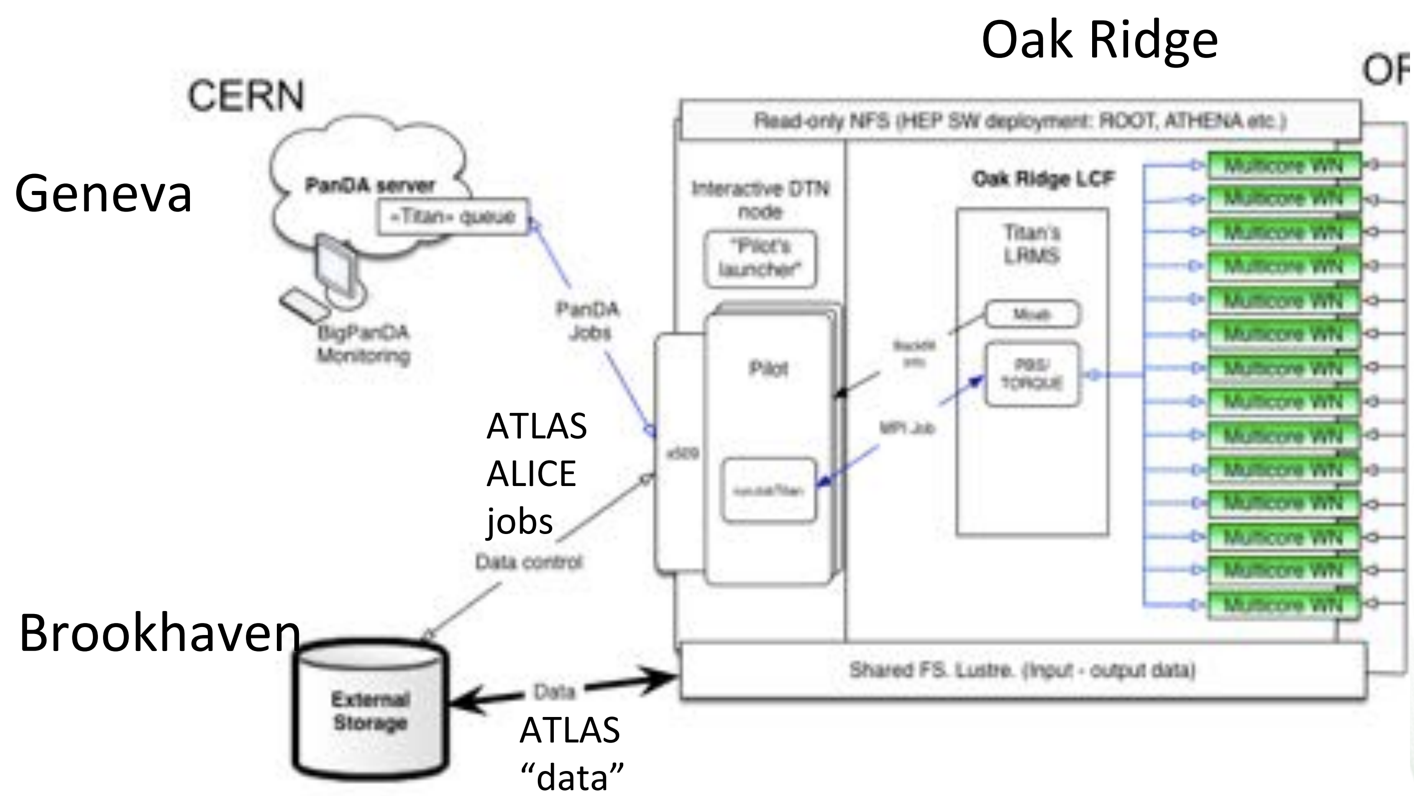
- Titan's architecture, configuration and policies poses several challenges to the deployment of PanDA e.g., :
  - The default deployment model of PanDA Pilot is unfeasible on Titan:
    - Pilot requires communication with the PanDA Server to pull jobs to execute, but
    - This not possible on Titan because worker nodes do not offer outbound network connectivity
  - The specific characteristics of the execution environment require re-engineering of ATLAS application frameworks, e.g., absence of local storage on the worker nodes, and modules tailored to Compute Node Linux.
- Given various constraints and challenges, the Monte Carlo detector simulation task is most suitable for execution on Titan at the moment.
  - Performed via AthenaMP, the ATLAS soft- ware framework integrating the GEANT4 detector simulation toolkit.
  - Accounts for ≈ 60% to 70% of all the jobs on WLCG, making them a primary candidate for offloading.
  - Task is mostly computational-intensive, requiring less than 2GB of RAM at runtime and small input data.

Deploying PanDA on Titan (2)

Given the communication requirements of PanDA Pilots and the unused capacity of Titan, PanDA pilot was repurposed to serve as a job broker on the DTN nodes of Titan.

- Maintaining the core modules of PanDA Pilot and its stand-alone architecture, this prototype called ‘PanDA Broker’ supports the execution of MC detector simulations.
- Each PanDA Broker queries the PanDA server for ATLAS jobs that have been bound to Titan.
- Upon receiving jobs descriptions, PanDA Broker pulls jobs’ input files from BNL to the OLCF Lustre file system.
- PanDA Broker queries Titan’s Moab scheduler about the current available backfill slot and creates an MPI script, wrapping enough ATLAS jobs’ payload to fit the backfill slot.
- PanDA Broker submits the MPI script to the Titan’s Torque batch system via RADICAL- SAGA.

OLCF Titan Integration with ATLAS Computing



D. Oleynik, S. Panitkin, M. Turilli, A. Angius, S. Oral, K. De, A. Klimentov, J. C. Wells and S. Jha, "High-Throughput Computing on High-Performance Platforms: A Case Study", IEEE e-Science (2017) available as: https://arxiv.org/abs/1704.00978

Backfill Optimization

- Backfill optimization of Titan's Moab scheduler allows to avoid the overhead of queue wait times without using the traditional pilot abstraction.
  - With this optimization, Moab starts low-priority jobs when they do not delay higher priority jobs, independent of whether the low-priority jobs were queued after the high-priority ones.
- With "showbf" command, users can interrogate Moab about the number of worker nodes and wall time immediately available.
- If a job is immediately submitted to Titan requesting that number of worker nodes and walltime, chances are that Moab will immediately schedule it.

Adaptive Computing Enterprises. (2014) Maui scheduler administrator's guide: Backfill. Available: http: //docs.adaptivecomputing.com/maui/8.2backfill.php

OLCF Job Priority by Node Count

Used for OLCF's Jaguar and Titan eras

- The OLCF implements queue policies that encourage the submission and timely execution of large, leadership-class jobs on Titan.
- Jobs are aged according to the job's requested processor count,
  - (older age equals higher queue priority).
- Each job's requested node count places it into a specific bin. Each bin has a different aging parameter, which all jobs in the bin receive.

Bin	Min Nodes	Max Nodes	Max Walltime (Hours)	Aging Boost (Days)
1	11,250	18,688	24	+15
2	3,750	11,249	24	+5
3	313	3,749	12	0
4	126	312	6	0
5	1	125	2	0

OLCF Allocation Overuse Policy

- Projects that overrun their allocation are still allowed to run on OLCF systems, although at a reduced priority.
- Like the adjustment for the number of processors requested above, this is an adjustment to the apparent submit time of the job.

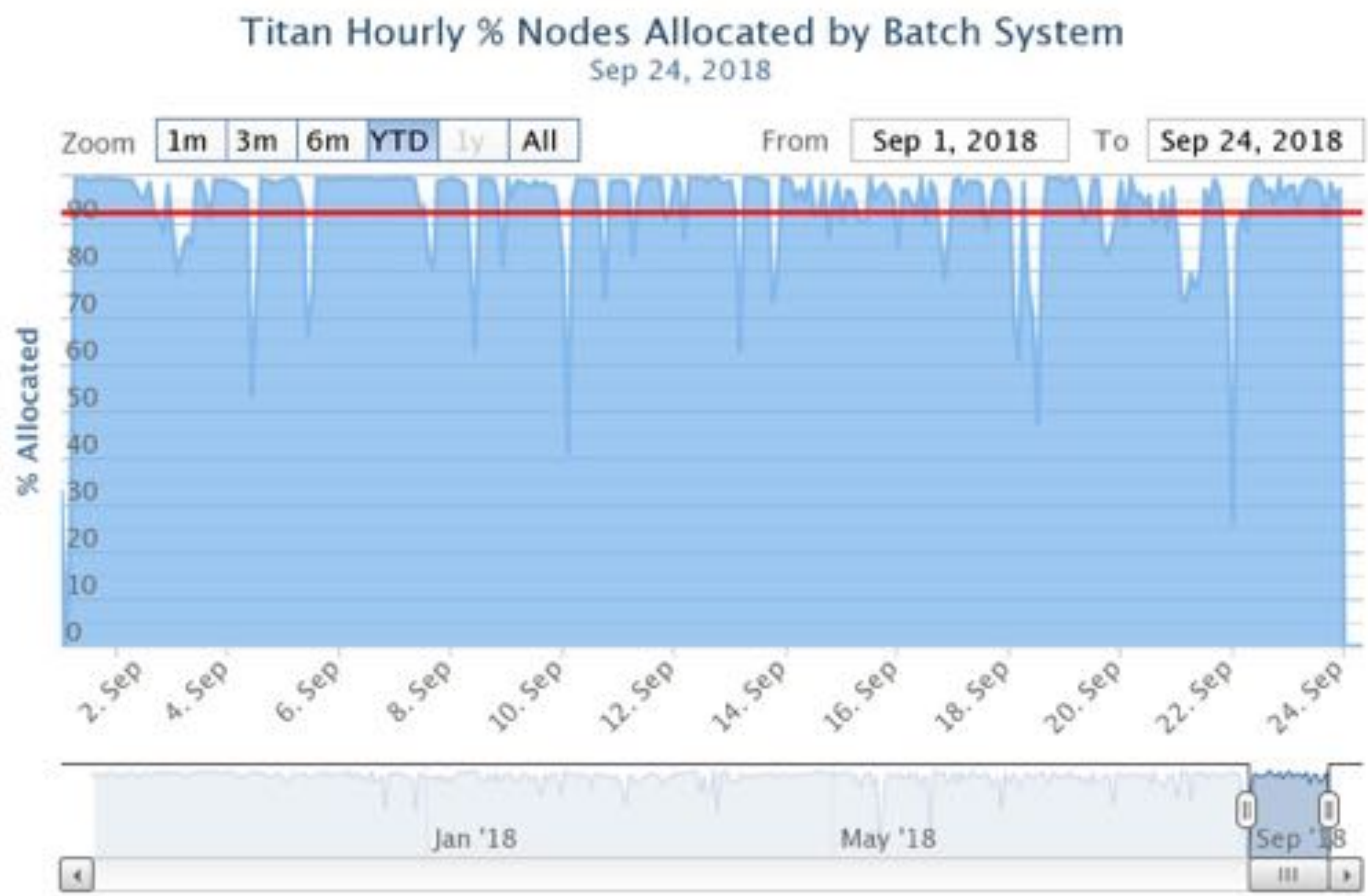
% Of Allocation Used	Priority Reduction	number eligible-to-run	number running
< 100%	0 days	4 jobs	unlimited jobs
100% to 125%	-30 days	4 jobs	unlimited jobs
> 125%	-365 days	4 jobs	1 job

Policies Evolved for BigPanDA Demo (CSC108)

Allocation	Priority Reduction	number eligible-to-run	number running
No Allocation	-366 days	20 jobs	20 jobs

- CSC108 jobs are always the lowest-priority jobs in Titan’s queue.
- Jobs are not allowed to age – terminated after 5 minutes if not scheduled.

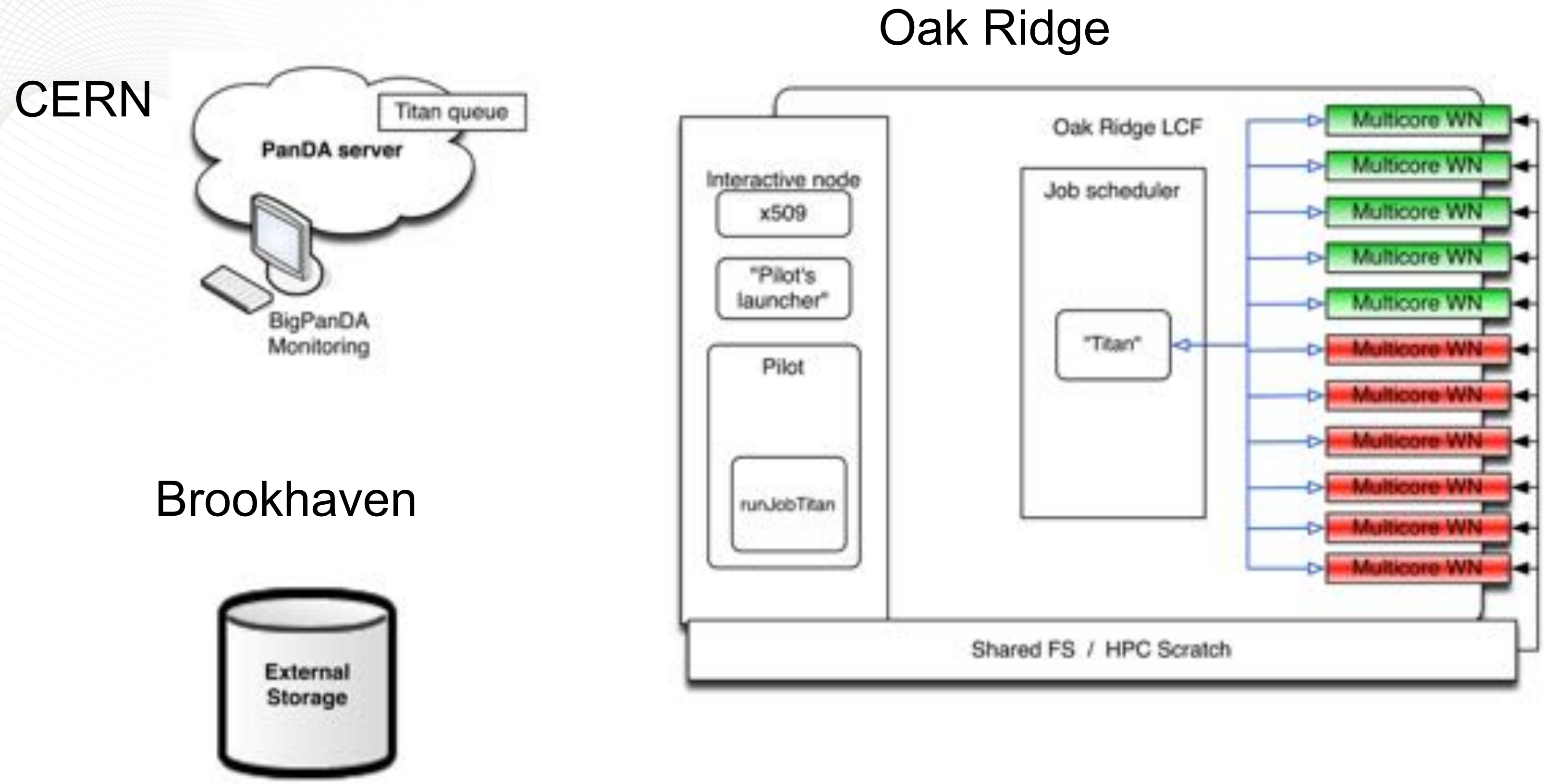
Utilization of Available Compute Resources on Titan: Sept. 2018



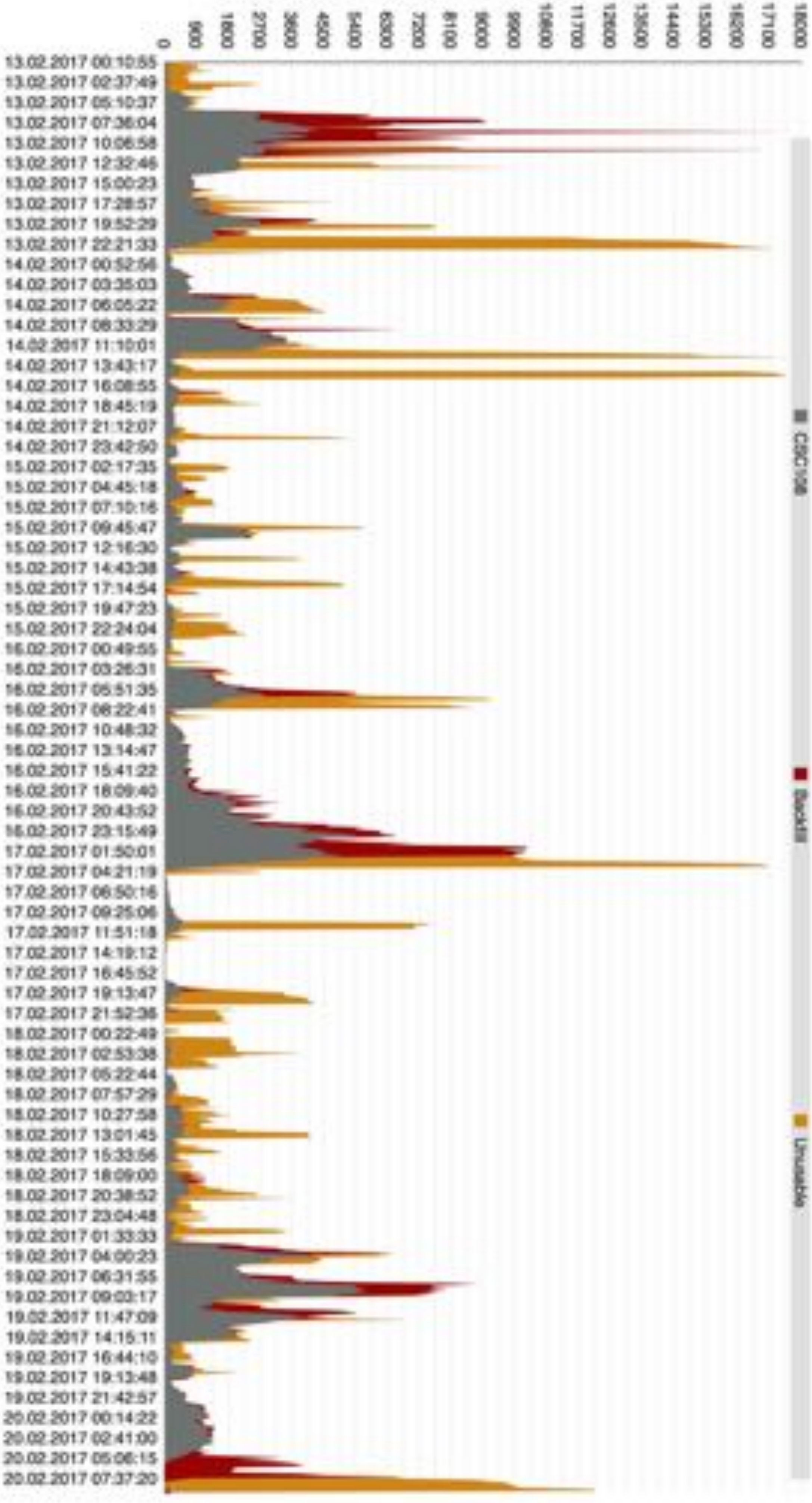
Titan Hourly % Nodes Allocated by Batch System: 1 September – 24 September, 2017

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

OLCF Titan Integration with ATLAS Computing



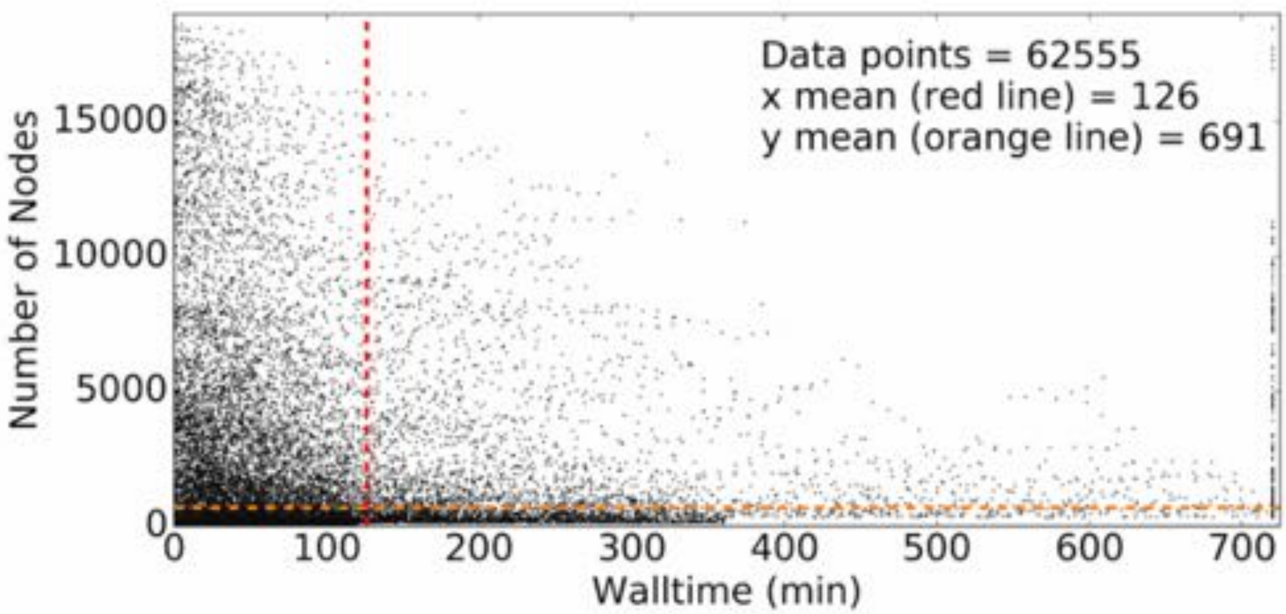
D. Oleynik, S. Panitkin, M. Turilli, A. Angius, S. Oral, K. De, A. Klimentov, J. C. Wells and S. Jha, "High-Throughput Computing on High-Performance Platforms: A Case Study", 2017 IEEE e-Science Conference, DOI: 10.1109/eScience.2017.43


0
900
1800
2700
3600
4500
5400
6300
7200
8100
9000
9900
10800
11700
12600
13500
14400
15300
16200
17100
18000
13.02.2017 00:10:55
13.02.2017 02:37:49
13.02.2017 05:10:37
13.02.2017 07:36:04
13.02.2017 10:06:58
13.02.2017 12:32:46
13.02.2017 15:00:23
13.02.2017 17:28:57
13.02.2017 19:52:29
13.02.2017 22:21:33
14.02.2017 00:52:56
14.02.2017 03:35:03
14.02.2017 06:05:22
14.02.2017 08:33:29
14.02.2017 11:10:01
14.02.2017 13:43:17
14.02.2017 16:08:55
14.02.2017 18:45:19
14.02.2017 21:12:07
14.02.2017 23:42:50
15.02.2017 02:17:35
15.02.2017 04:45:18
15.02.2017 07:10:16
15.02.2017 09:45:47
15.02.2017 12:16:30
15.02.2017 14:43:38
15.02.2017 17:14:54
15.02.2017 19:47:23
15.02.2017 22:24:04
16.02.2017 00:49:55
16.02.2017 03:26:31
16.02.2017 05:51:35
16.02.2017 08:22:41
16.02.2017 10:48:32
16.02.2017 13:14:47
16.02.2017 15:41:22
16.02.2017 18:09:40
16.02.2017 20:43:52
16.02.2017 23:15:49
17.02.2017 01:50:01
17.02.2017 04:21:19
17.02.2017 06:50:16
17.02.2017 09:25:06
17.02.2017 11:51:18
17.02.2017 14:19:12
17.02.2017 16:45:52
17.02.2017 19:13:47
17.02.2017 21:52:36
18.02.2017 00:22:49
18.02.2017 02:53:38
18.02.2017 05:22:44
18.02.2017 07:57:29
18.02.2017 10:27:58
18.02.2017 13:01:45
18.02.2017 15:33:56
18.02.2017 18:09:00
18.02.2017 20:38:52
18.02.2017 23:04:48
19.02.2017 01:33:33
19.02.2017 04:00:23
19.02.2017 06:31:55
19.02.2017 09:03:17
19.02.2017 11:47:09
19.02.2017 14:15:11
19.02.2017 16:44:10
19.02.2017 19:13:48
19.02.2017 21:42:57
20.02.2017 00:14:22
20.02.2017 02:41:00
20.02.2017 05:06:15
20.02.2017 07:37:20
CSC108
Backfill
Unusable

Operational Demo: Scaling Up Active Backfill

- Consumed 370 Million Titan core hours from January 2016 to present
    - This is 2.8 percent of total available time on Titan over this period
- Remaining used backfill slots are often too short or too small for assigned ATLAS payloads

Understanding Backfill Slot Availability

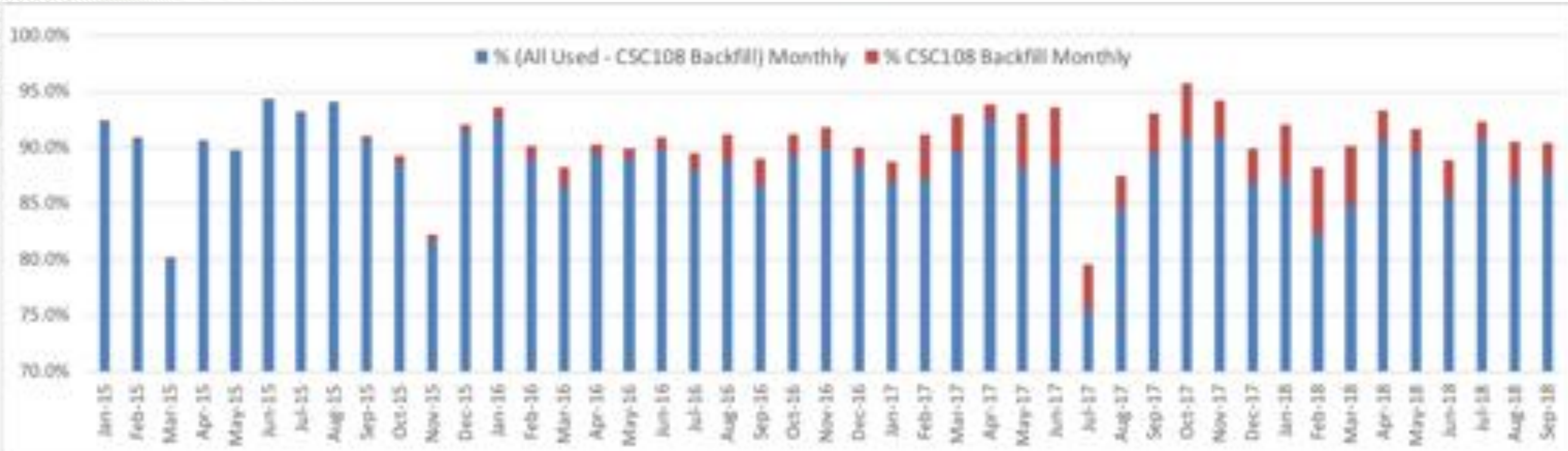


2555 measures of Backfill availability on Titan during a time window.

- Mean Backfill availability: 691 worker nodes for 126 minutes.
- Up to 15K nodes for 30-100 minutes
- Large margin for optimization

D. Oleynik, S. Panitkin, M. Turilli, A. Angius, S. Oral, K. De, A. Klimentov, J. C. Wells and S. Jha, ”High-Throughput Computing on High-Performance Platforms: A Case Study”, 2017 IEEE e-Science Conference, DOI: 10.1109/eScience.2017.43

Operational Demo: Scaling Up Active Backfill

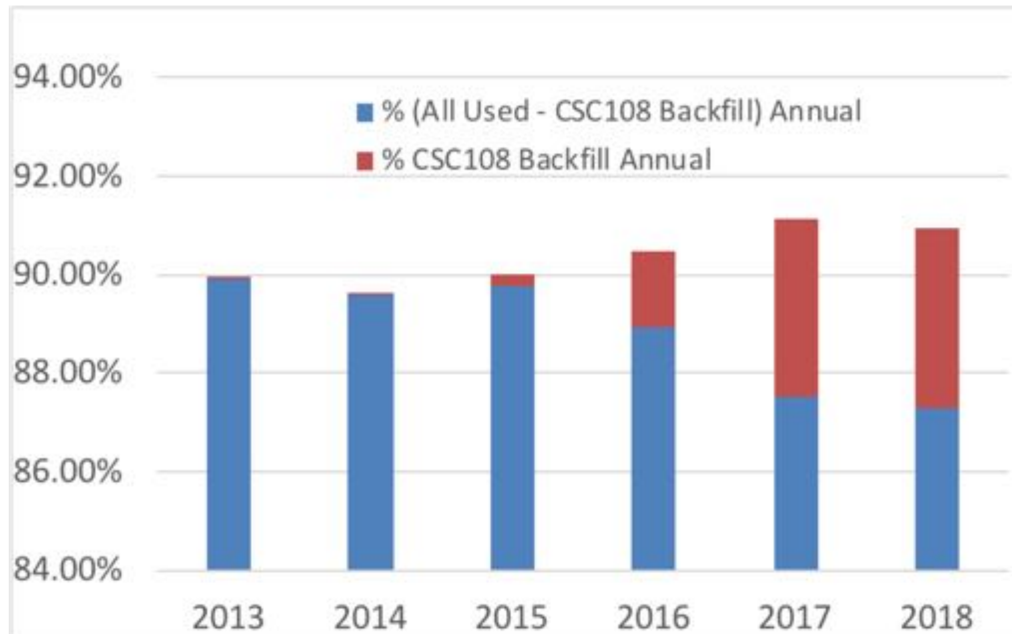


- Slowly increased over time number of PanDA brokers from 1, 4, 8, 16, 20
- Increase the maximum number of nodes per broker to 350 nodes.

Operational Demo: Scaling Up Active Backfill

- Increased Titan's utilization by ~2 percent over historical trends
  - Have we displaced ~ 2 percent of Titan's traditional throughput?
  - Negative impact on throughput of traditional work loads is not obvious, under evaluation.
- Preemption of PanDA payloads to be implemented.

BigPanDA Backfill as a Percentage of overall Titan Utilization



BigPanDA demo contributed to winning ALCC project

2017 ASCR Leadership Computing Challenge (ALCC) Application

Assigned Proposal ID:

286

Title:

2017 ASCR Leadership Computing Challenge (ALCC) Application

Consortium/End-Station Proposal

Principal Investigator:

John T Childers, Argonne National Laboratory, Tel: 3313024647. Email: jchilders@anl.gov

Project collaborators:

Thomas LeCompte (Argonne National Laboratory)
Doug Benjamin (Duke University)
Radja Boughezal (Argonne National Laboratory)
Paolo Calafiura (Lawrence Berkeley National Laboratory)
Stefan Hoeche (SLAC National Laboratory)
Burt Holzman (Fermi National Accelerator Laboratory)
Alexei Klimentov (Brookhaven National Laboratory)
Jim Kowalkowski (Fermi National Accelerator Laboratory)
Frank Petriello (Northwestern University)
Vakho Tsulaia (Lawrence Berkeley National Laboratory)
Craig Tull (Lawrence Berkeley National Laboratory)
Thomas Uram (Argonne National Laboratory)
Torre Wenaus (Brookhaven National Laboratory)

HPC resources requested:

Titan: 80 x10^6 Titan-core hours, ~20 TB online storage, ~0 TB offline storage

Mira: 18 x10^6 Mira-core hours, ~20 TB online storage, ~0 TB offline storage

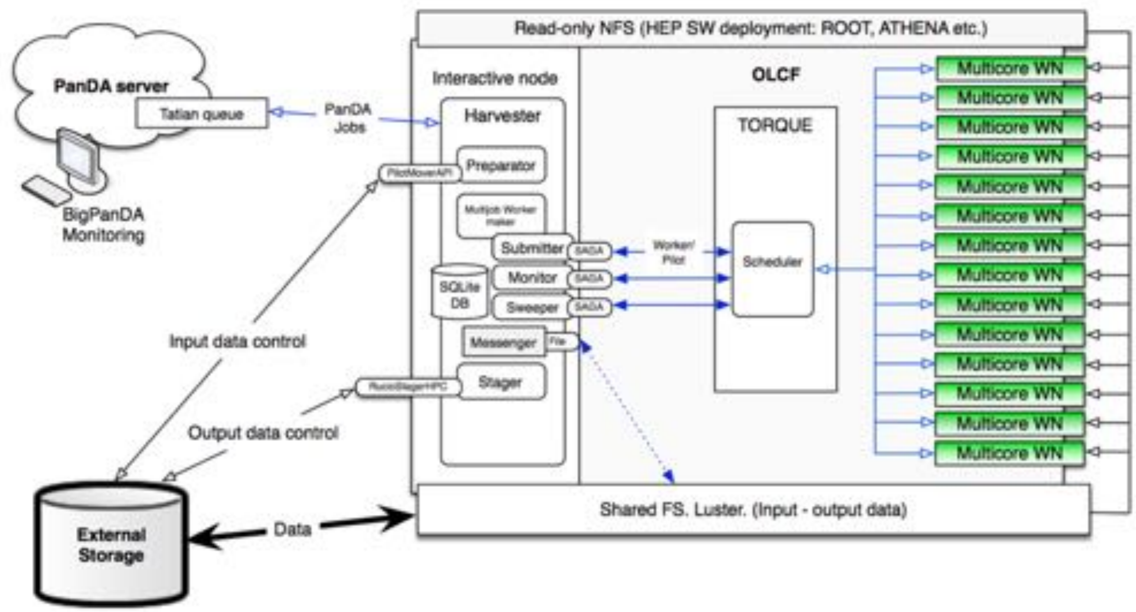
Theta: 45.5 x10^6 Theta-core hours, ~20 TB online storage, ~0 TB offline storage

Cori (Cray XC40 Intel Xeon Phi KNL nodes): 36 x10^6 NERSC-core hours, ~100 TB scratch storage, ~10 TB proje

Cori/Edison (Cray XC40/30 Intel Xeon nodes): 33 x10^6 NERSC-core hours, ~100 TB scratch storage, ~10 TB proj

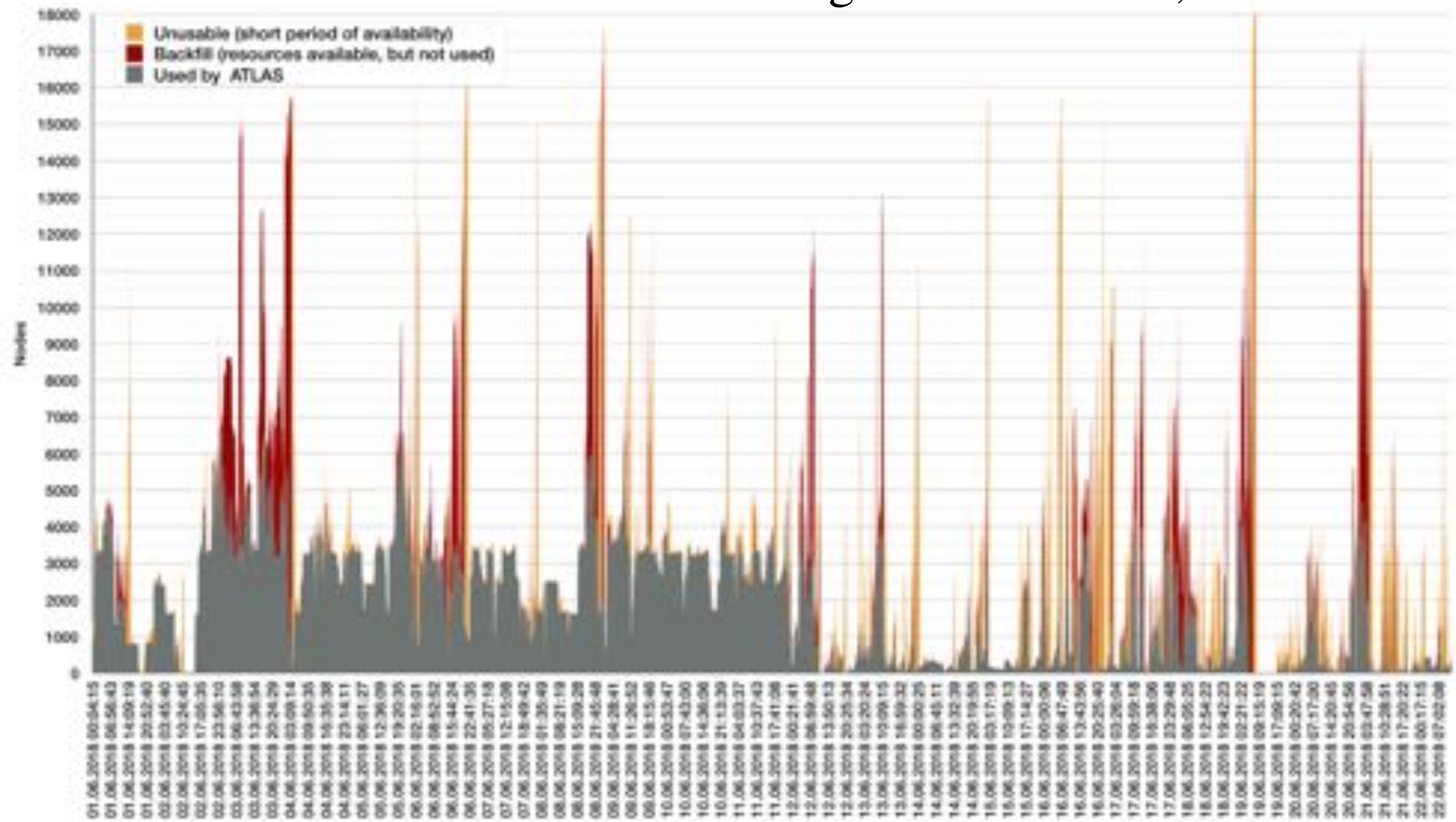
Additional notes: At NERSC, the request for Storage is in total. That means 100TB (Scratch) + 10TB (Project) across both Cori and Cori/Edison. Ideally these two spaces would be visible to both machines.

Implement ALCC Project at OLCF using Harvester

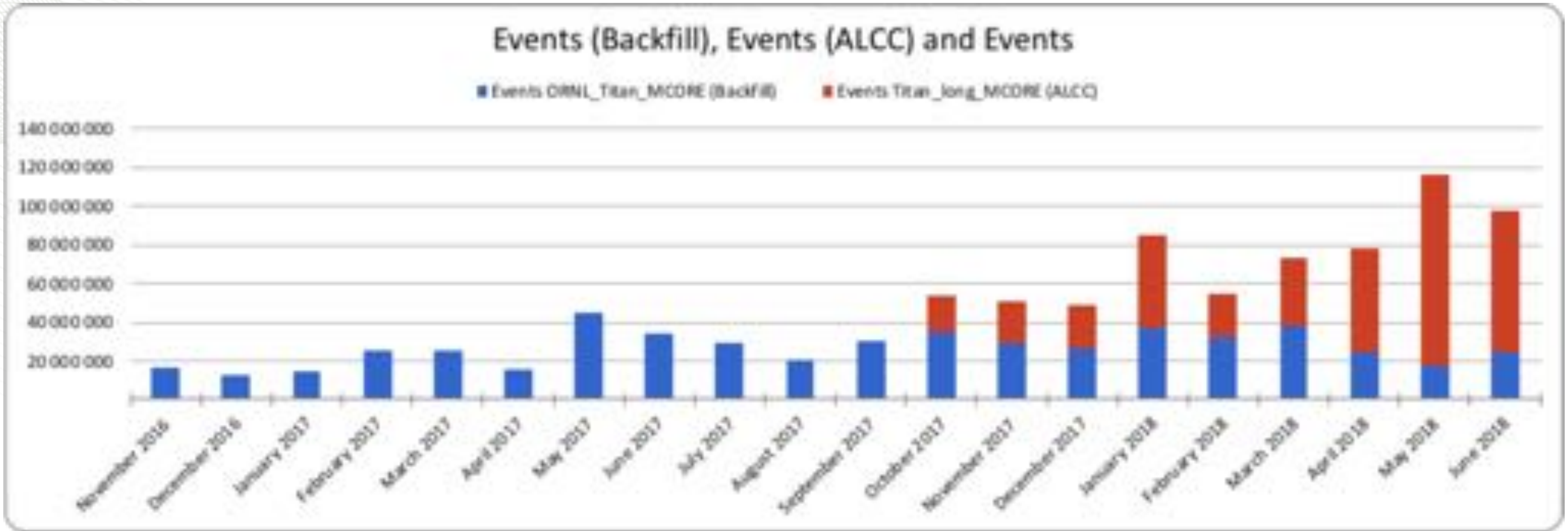


ATLAS@OLCF: Batch Queue Submission & Active Backfill

- Backfill utilization in 1 June through 22 June 2018, 10-min data frequency



Plot:
Danila Oleynik



- Since Nov. 2016, 508 million ATLAS events computed via backfill
- Since Oct. 2017, 395 million ATLAS events computed via "normal" batch queue
  - Increases in batch queue event generation beginning in Feb. 2018 show the impact of Harvester

Broader HPC Impacts

- While PanDA Broker implementation is resource specific, it was successfully ported to other supercomputers, including the
  - Edison/Cori at the National Energy Research Scientific Computing Center (NERSC),
  - Piz Daint, (CSCS)
  - MareNostrum, (BSC)
  - IT4I, Czech Republic
  - SuperMUC, Munich

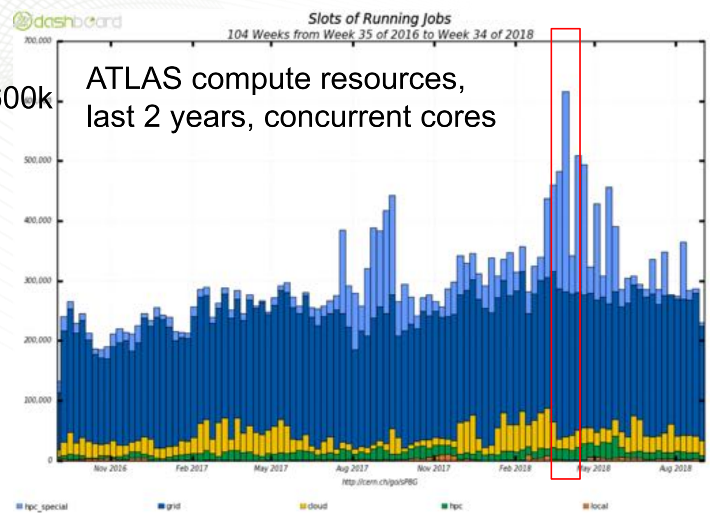
- Significant impact on HEP planning for computing resources.

The role of HPC for HEP computing - Situation and Outlook

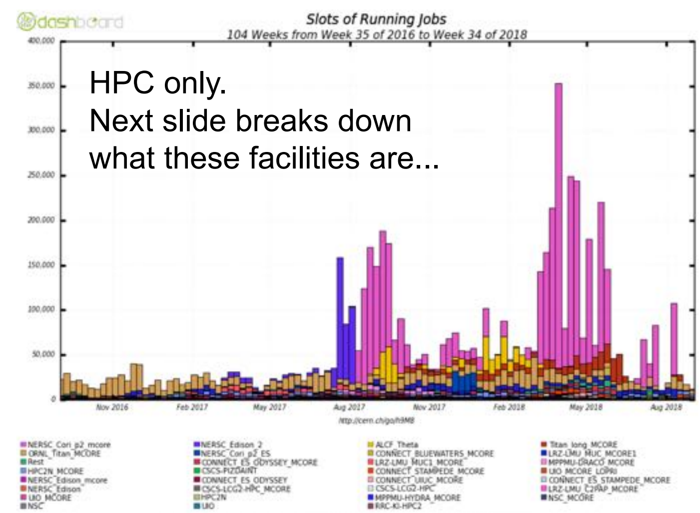
Torre Wenaus (BNL)

Scientific Computing Forum Meeting
CERN
September 20 2018

HPC usage in ATLAS



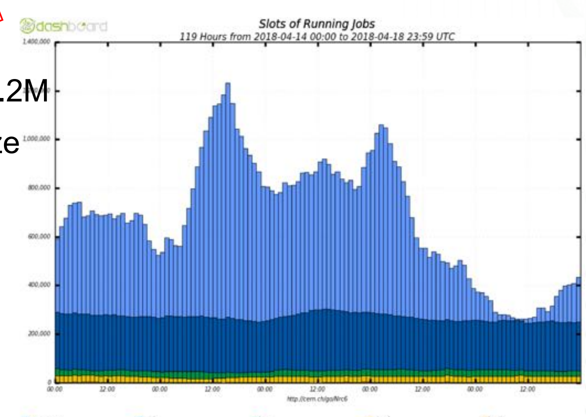
A long history but a new era in the last year: very large facilities, so far in the US



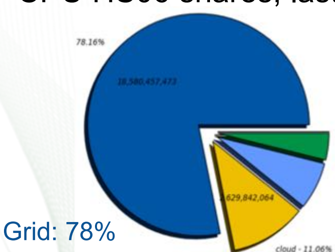
Light blue: “special” HPCs, where special means big, difficult to use, US DOE
Dark blue: the grid
Yellow: cloud resources including (dominantly) HLT
Green: “regular” HPCs, meaning easier to use, European or US NSF

Zoom showing full size of scaling peak: 1.2M concurrent cores.

Our workload management system is highly scalable!



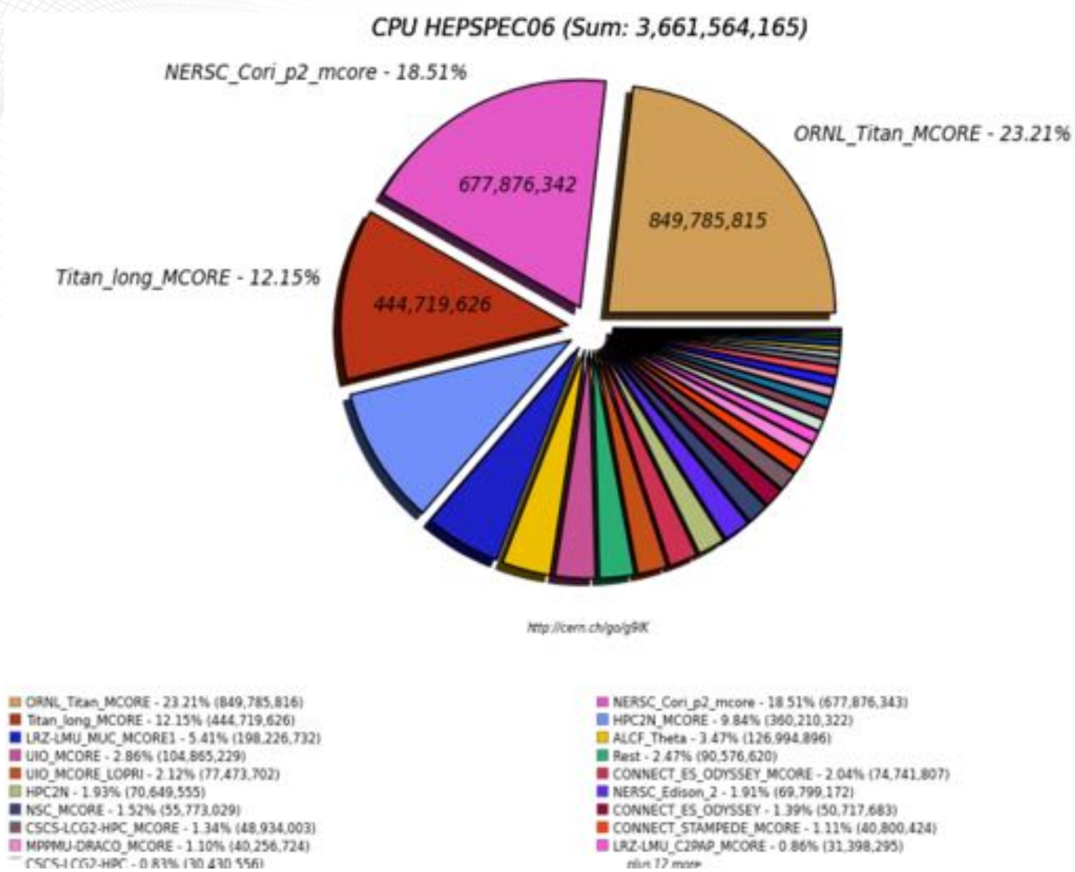
CPU HS06 shares, last year



Grid: 78%
Cloud, HLT: 11%
HPC special: 7%
HPC regular: 4%

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

HPCs in ATLAS



Breakdown of the HPC facilities in the previous plots

- US DOE HPCs (all in the “special” category)
  - Titan at Oak Ridge
  - Cori at NERSC (successor to Edison)
  - Theta at ANL (successor to Mira)
- Nordugrid
  - Several of their facilities are HPCs, including HPC2N #4
- European HPCs
  - LRZ (SuperMUC), MPPMU, CSCS, ...
- US NSF HPCs
  - Sites with ‘CONNECT’ in their name

All in routine production, mostly Geant4 MC simulation

HPCs in HEP: US DOE view

Similar views from HEPAP panel (supplementary slide)

What We've Learned So Far

- HPC architectures will continue to evolve, but moving to vectorized, multithreaded codes tailored to I/O-bound systems will result in higher efficiency codes
  - Engaging HPC experts to analyze code has helped identify algorithm alternatives and data flow bottlenecks, in some cases resulting in spectacular speedups (e.g. 600x). Continued engagement is therefore essential!
  - Need to identify which codes could benefit the most
- Using Exascale machines badly (e.g. by ignoring the GPU/accelerator) will result in a factor-of-40 penalty in performance that will not be tolerated. HEP will lose its allocations if it does this.
  - Engaging Exascale Computing Project (ECP) experts early and often will result in faster adoption of best practices for exascale machines, and influence ECP design choices to HEP's benefit. HEP needs a coordinated interface to both ECP & the Leadership Computing Facilities.
  - Need to identify which codes could benefit the most
- LQCD regularly rewrites its code, has reaped significant speedup benefits every time
- Reinforced that multiyear NERSC allocations & better metrics for pledges are needed
- End-to-end network data flow models are needed to support tradeoff analysis of storage vs. CPU vs. network bandwidth on a system-wide and program-wide basis
  - Greater sharing of the underlying data management software layer may also be beneficial

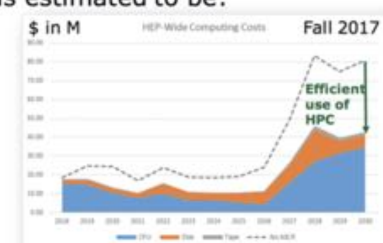
U.S. DEPARTMENT OF ENERGY | Office of Science — DOE HEP Status at HEPAP - May 2018 — 28

We must use them properly (use the accelerators)

We must use them heavily

Updated HEP Computing Model

- In preparation for the Inventory Roundtable, the largest HEP experiments from all three frontiers were asked to provide a **more detailed estimate** of their expected computing needs
  - CPU, storage, network, personnel, and HPC portability
- Cost estimates for all experimental frontiers:
  - "Business as usual" (minimal additional HPC use): **$600M ± 150M**
  - With effective use of HPC resources this reduces to: **$275M ± 70M**
- By 2030 cost share by frontier is estimated to be:
  - ½ Energy Frontier
  - ¼ Intensity Frontier
  - ¼ Cosmic Frontier
- **A strategy encompassing all HEP computing needs is required!**

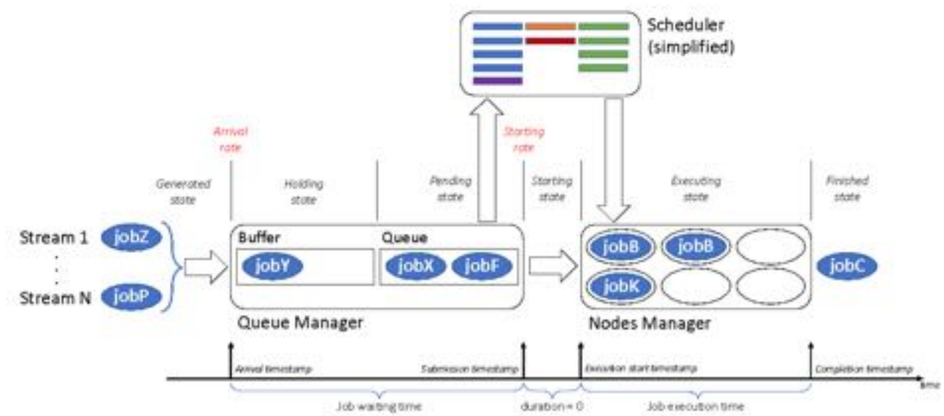


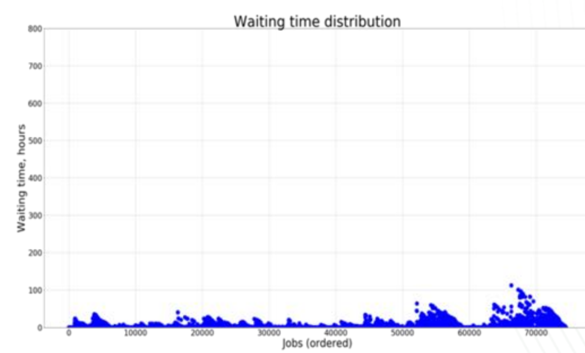
U.S. DEPARTMENT OF ENERGY | Office of Science — DOE HEP Status at HEPAP - May 2018 — 29

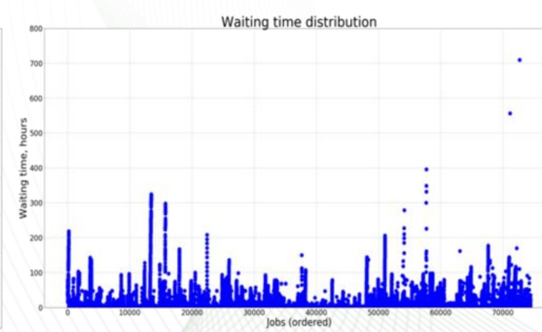
Jim Siegrist, HEPAP meeting, May 2018

Modelling Resource Utilization on Titan

- Aim: To provide insight into adaptive execution parameters and "optimal" way to configure jobs (backfill and regular q) to increase resource utilization
- Two distinct modelling strategies:
  - Simulator: Model the load (= number of core-hours used) given a job trace
  - Probabilistic Model: Estimate the number of core-hours utilized for background distribution of job size and waiting time



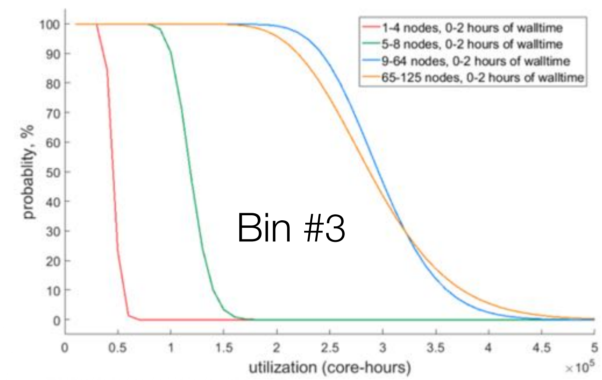


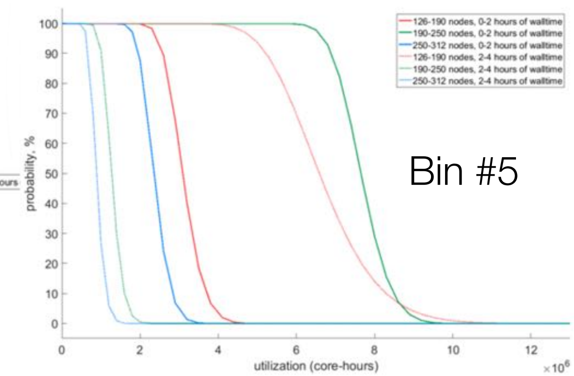


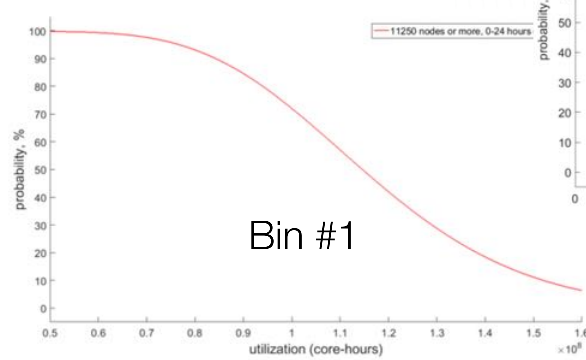
OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Modelling Resource Utilization on Titan

- Aim: To provide insight into adaptive execution parameters and "optimal" way to configure jobs (backfill and regular q) to increase resource utilization
- Two distinct modelling strategies:
  - Simulator: Model the load (= number of core-hours used) given a job trace
  - Probabilistic Model: Estimate the number of core-hours utilized for background distribution of job size and waiting time







Bin #4

OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

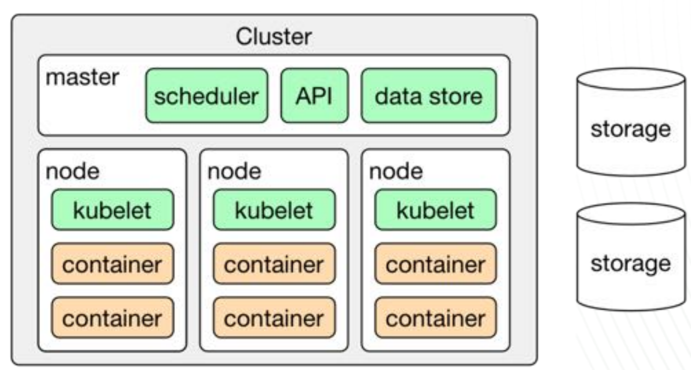
Basic HPC Workflow Requirements at OLCF

- Need ways for users to manage their workflow system
  - Diverse ecosystem of workflow systems makes it difficult for NCCS Operations to support every one
- Why not just use SSH keys?
  - Our moderate security controls require remote actions to be authenticated with RSA two-factor credentials
  - Instead we are working on providing ability for running workflow services locally
- Upon surveying existing workflow systems we came up with the following requirements:
  - Run a persistent service locally as a “daemon” that stays up
  - Talk to batch submission system for current queue information and job submission
  - Interact with files on GPFS/Lustre/NFS

OpenShift

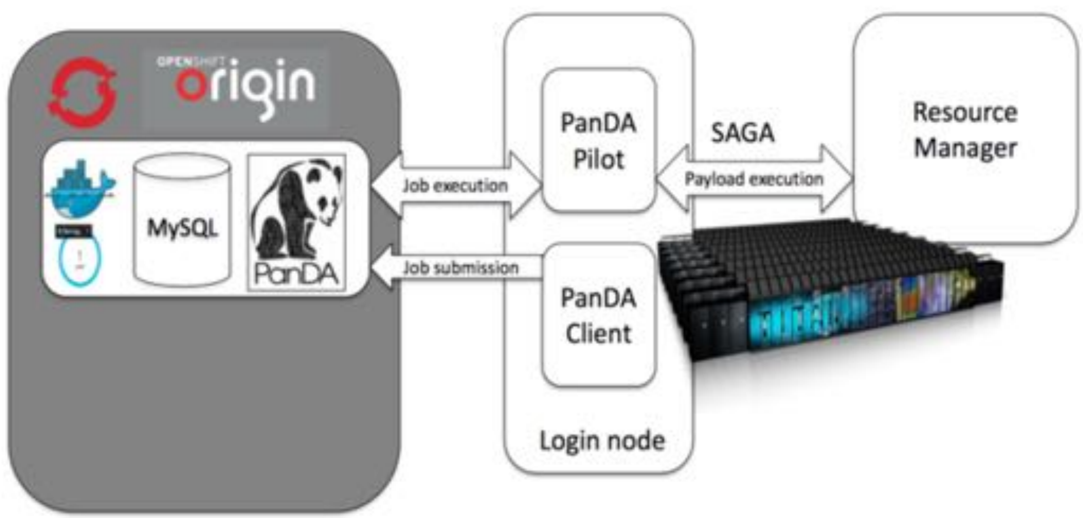
RED HAT® OPENSHIFT Container Platform

- Distribution of Kubernetes developed by Red Hat
- Kubernetes manages containerized applications across nodes and provides mechanisms for deployment, maintenance, and application-scaling.
- Analogous to but separate from a batch scheduler on a compute cluster



PanDA Server at OLCF: Broad application across domains

- In March 2017 a new PanDA server instance has been set up at ORNL to serve various experiments. This installation the first at OLCF to demonstrate application of a container cluster management and orchestration system, Red Hat OpenShift Origin.
  - We are looking forward for experimenting with a wide variety of payloads.
- OpenShift, when fully in production, will give OLCF users the ability to deploy and manage infrastructure services
  - https://www.olcf.ornl.gov/2017/06/05/olcf-testing-new-platform-for-scientific-workflows/



OPENSHIFT

Key Contributors:
Jason Kincl (ORNL),
Ruslan.Mashinistov (BNL)

PanDA Server at OLCF: PanDA WMS beyond HEP

- Biology / Genomics: Center for Bioenergy Innovation at ORNL
- Molecular Dynamics: Prof. K. Nam (U. Texas-Arlington)
- IceCube Experiment
- Blue Brain Project (BBP), EPFL
- LSST (Large Synoptic Survey Telescope) project
- LQCD, US QCD Project
- nEDM, (neutron Electric Dipole Moment Experiment, ORNL

Conclusions

- DOE Office of Science is advancing an integrated vision for exascale computing ecosystem, including data-intensive applications, e.g., experimental and observational data.
- Big PanDA deployment of Titan shows the potential of distributed high-throughput computing to be integrated with high-performance computing infrastructure.
  - Offers significate value for other projects beyond HEP
- Solving this problem of HPC-HTC integration, even in this "proof-of-concept" demonstration, has opened new vistas for HENP computing.
- Big PanDA project provided the initial motivation for the development of OLCF's strategy for container orchestration for user-specified HPC middleware services.

BROOKHAVEN NATIONAL LABORATORY
UNIVERSITY OF TEXAS ARLINGTON
THE STATE UNIVERSITY OF NEW JERSEY RUTGERS
OAK RIDGE National Laboratory | LEADERSHIP COMPUTING FACILITY

Questions? Jack Wells, wellsjc@ornl.gov