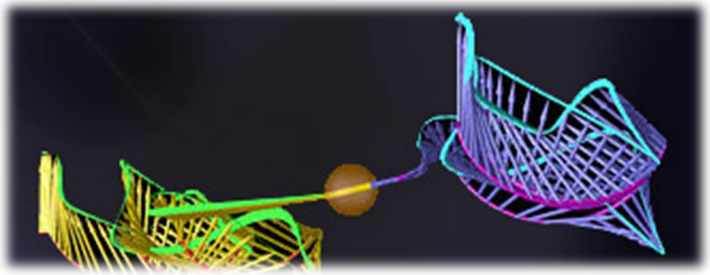


LQCD-ext II
2019 Accomplishments & Performance

Bill Boroski
LQCD Contractor Project Manager

DOE Scientific Review
Rockville, Maryland
July 9-10, 2019

Outline

- Response to 2018 annual review suggestions
- LQCD-ext II project scope, budget, and organization
- FY19 hardware acquisition activities
- FY19 technical and financial performance
- Summary

Response to Suggestions from the DOE 2018 Annual Review

Response to 2018 Annual Review Suggestions (1 of 6)

- Suggestion #1: Each of the reviewers recommended that the remaining HEP FY2018 project funds ($0.85m) be released as soon as possible in light of the very positive impressions made at the review.

  - *Response: Following the review, the Office of High Energy Physics (HEP) released the remaining $0.85m of FY18 funds to Fermilab, the site of the next planned acquisition. All remaining funds were received at Fermilab by the end of July 2018.*

Response to 2018 Annual Review Suggestions (2 of 6)

- Suggestion #2: The project should work with Fermilab to initiate the development of a program of Institutional Clusters.
  - *Response: The LQCD Project and Fermilab collaborated on the design, procurement and installation of a high-performance computing cluster that met the computing needs of LQCD and the Fermilab scientific community.*
    - *A Joint Acquisition Planning Committee was formed with members chosen based on their technical expertise and knowledge of USQCD needs, Fermilab experimental program needs, or both.*
    - *The committee chairperson was Amitoj Singh, an FNAL employee and FNAL Site Architect for the LQCD Project.*
    - *The purpose of the committee was to understand user needs and existing computing resources, and to make a recommendation on the design and specifications for a new institutional computing cluster at Fermilab.*
    - *The committee issued a written report with recommendation in September 2018.*

Response to 2018 Annual Review Suggestions (3 of 6)

- Suggestion #3: The project should develop procedures to document scientific milestones uniformly over all the LQCD areas so that the project can track their annual progress quantitatively and present it more thoroughly at each review.

  - *Response: USQCD commissioned six whitepapers on the full range of physics topics, and a seventh on computing accomplishments and challenges. The authors represented all our scientific goals and matched physics relevance and computing feasibility into a set of reviewable milestones.*

Response to 2018 Annual Review Suggestions (4 of 6)

- Suggestion #4: Given the direct relevance of lattice gauge calculations to the experimental community, it would be valuable to enlist experimental physicists to advocate for the project during future reviews and in the next multi-year extension proposal due in 2019.

  - *Response: We agree with this suggestion. Historically, both experimentalists and phenomenologists have joined the USQCD Scientific Advisory Board quite eagerly. Many of them have proven that they could be very useful to USQCD and DOE in this way, and we plan to draw on them for this purpose.*

Response to 2018 Annual Review Suggestions (5 of 6)

- Suggestion #5: The project team should formulate a written plan to address the decreasing satisfaction articulated in the Compute Facility Satisfaction user survey results and present it to the DOE within two months.
  - *Response: We did not produce and deliver a written plan. Rather, we addressed the shortcomings directly.*
    - *The Associate Contractor Project Manager (ACPM) documented the user survey results and feedback by category, and by site.*
    - *Fermilab responded to the survey results by updating online user documentation.*
    - *Communication was sent to the individuals who participated in the Survey stating "The FNAL Online user documentation has been updated to reflect the major changes" and the URL was supplied. http://www.usqcd.org/fnal*
    - *BNL created an USQCD introduction web page, updated online documentation that pertained to the institutional clusters, and implemented an improved process for handling service requests that uses a centralized online BNL system.*
    - *JLab action items are being tracked under the auspices of the Nuclear and Particle Physics LQCD Computing Initiative (NPPLCI)*

Response to 2018 Annual Review Suggestions (6 of 6)

- Suggestion #6: The reviewers recommend that Fermilab carefully examine the BNL institutional cluster model. The reviewers believe that Fermilab would discover that the advantages outweigh the disadvantages. It may even be beneficial for the laboratories to coordinate: do both Labs need to have the exact same mix of Single-CPU, multicore and GPU based computing?

  - *Response: Fermilab is implementing an Institutional Cluster system that follows the BNL institutional cluster model.*
    - *The planning process used for new acquisitions at both laboratories is very similar, particularly when the acquisition has LQCD impact.*
    - *Both laboratories are using the Slurm Workload Manager with similar-looking monitoring dashboards.*
    - *The laboratories coordinated with one another for the FY18 and FY19 acquisitions. Hardware decisions factored in the requirements of the LQCD project and the scientific programs at each laboratory.*
    - *When making hardware decisions, the needs of the USQCD user group are considered against the full LQCD hardware portfolio. Hardware configurations at the two laboratories are notably distinct.*

FY19 Progress Update

LQCD-ext II Progress Update

Original Scope

- Acquire and operate dedicated hardware at BNL, JLab, and FNAL for the study of QCD during the period FY2010-2014.
- Scope included acquisition, deployment, and operation of computing facilities; software development is out of scope.

Major project restructuring in FY2018 impacted project scope and mode of operations

- In January 2018, DOE Office of Nuclear Physics (NP) announced its intent to establish an NP-funded dedicated hardware project at JLab (*Nuclear and Particle Physics LQCD Computing Initiative, or* NPPLCI). Operations under the new structure began in February 2018.
- HEP-funded LQCD-ext II project transitioned from a dedicated hardware model (acquiring and operating dedicated hardware) to an institutional cluster (IC) operating model (purchasing node-hrs on institutional clusters operated and managed by the host laboratories).
- Whereas dedicated hardware systems were designed specifically to meet LQCD computing requirements, more "general purpose" institutional clusters are designed to meet the needs of a broader user community. Fortuitously, architecture of existing systems at BNL (IC and KNL) satisfied LQCD needs; and LQCD engaged heavily in design of new IC systems.

LQCD-ext II Progress Update (2)

Current Status

- We're in the fifth year of the current 5-year project (Oct 2014-Sep 2019)
- Operations continue to run smoothly at each host laboratory
- LQCD-ext II project currently consists of deployments and operations at BNL and FNAL and is funded entirely by DOE Office of High Energy Physics (HEP)
- We continue to maintain information-sharing and knowledge transfer between the three host sites (bi-weekly Site Manager Meetings and quarterly DOE calls). Resources from all sites are available to all of USQCD.
- Between LQCD-ext II and the new NP initiative at JLab, we have received the full $14M of planned funding in accordance with the approved baseline plan, dated Oct 1, 2014

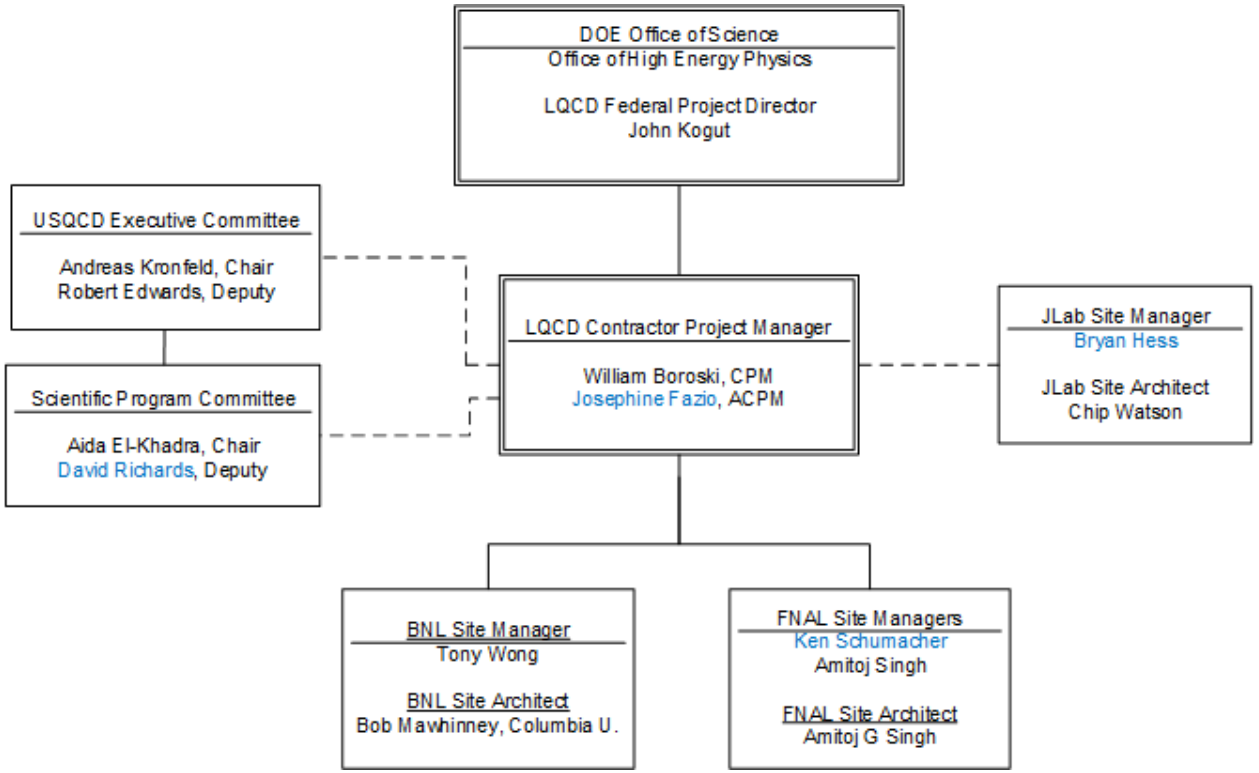
Funding Profile

Entity	FY15 (HEP & NP)	FY16 (HEP & NP)	FY17 (HEP & NP)	FY18	FY19	Total
LQCD-ext II	$2M	$3M	$3M	$2M (HEP)	$2M (HEP)	$12M
NPPLCI	---	---	---	$1M (NP)	$1M (NP)	$2M

LQCD-ext II Integrated Project Team (IPT)

Organizational changes since last year:

- Jo Fazio
- David Richards
- Ken Schumacher
- Bryan Hess



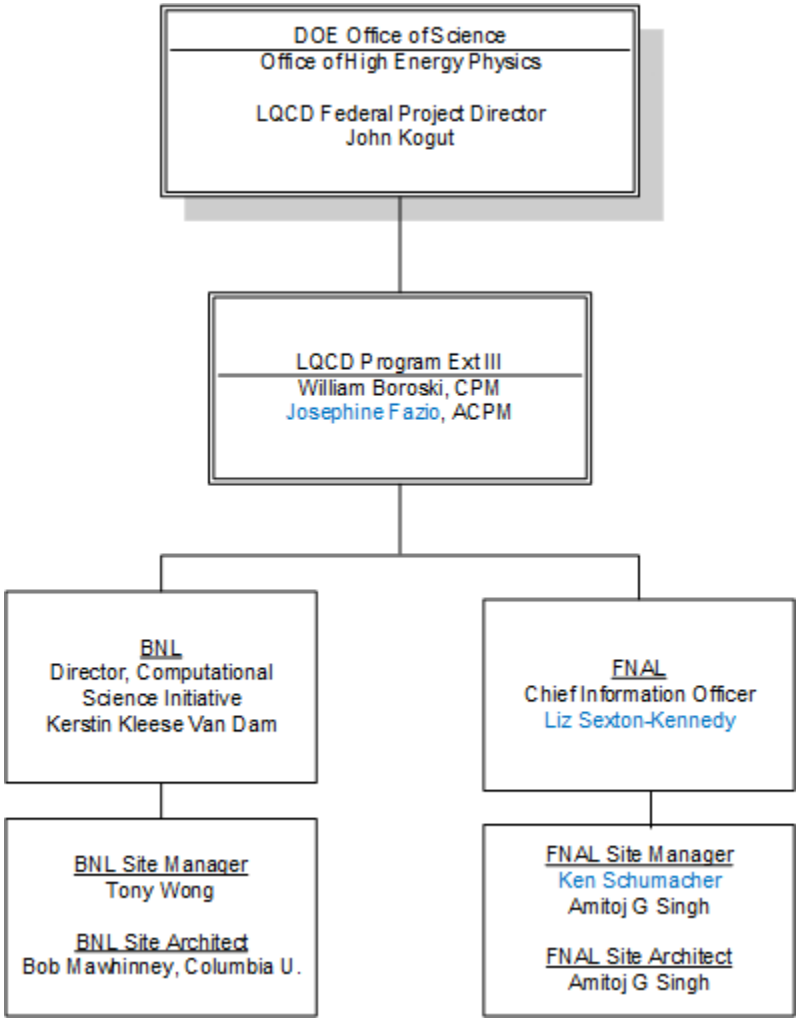
Formal Communication Channels:

- Bi-weekly Site Manager Calls (site managers & architects, project office, USQCD EC chairperson)
- Monthly DOE Progress Calls (DOE Project Director, project office, USQCD leadership)
- Annual USQCD All-Hands Meeting (Project office, site managers & architects, USQCD leadership & community)

LQCD-ext II Integrated Management Team

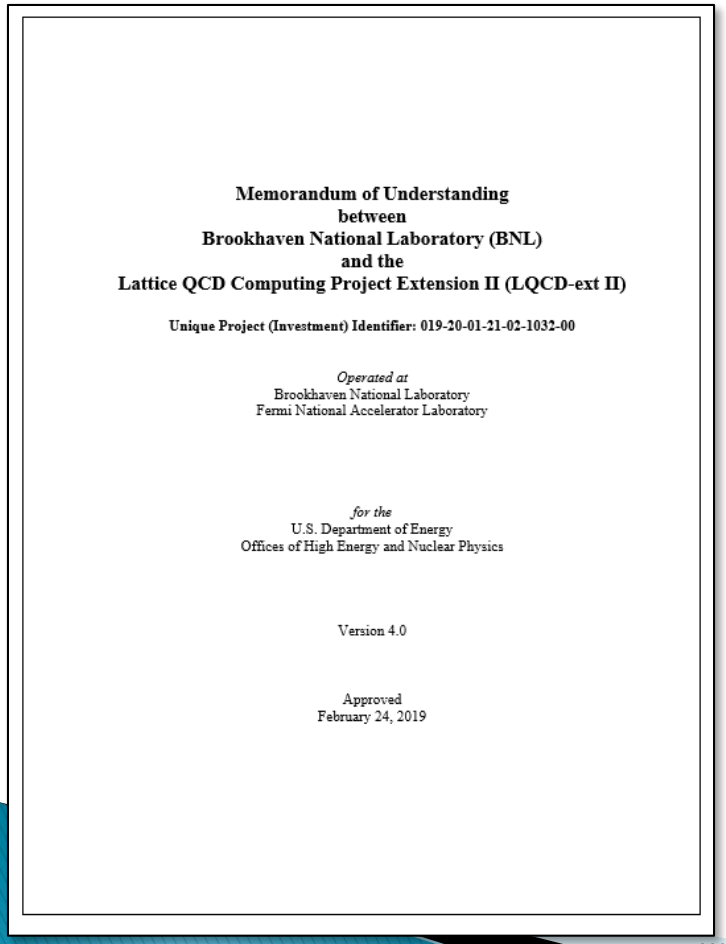
Organizational changes since last year:

- Liz Sexton-Kennedy
- Jo Fazio
- Ken Schumacher



Memoranda of Understanding (MOUs)

- MOUs have been executed that establish the commitments between the project and host labs, including node and storage allocations, per-node-hour and storage unit costs, reporting requirements, planned budget, etc.



Memorandum of Understanding
between
Brookhaven National Laboratory (BNL)
and the
Lattice QCD Computing Project Extension II (LQCD-ext II)

Unique Project (Investment) Identifier: 019-20-01-21-02-1032-00

Operated at
Brookhaven National Laboratory
Fermi National Accelerator Laboratory

for the
U.S. Department of Energy
Offices of High Energy and Nuclear Physics

Version 4.0

Approved
February 24, 2019

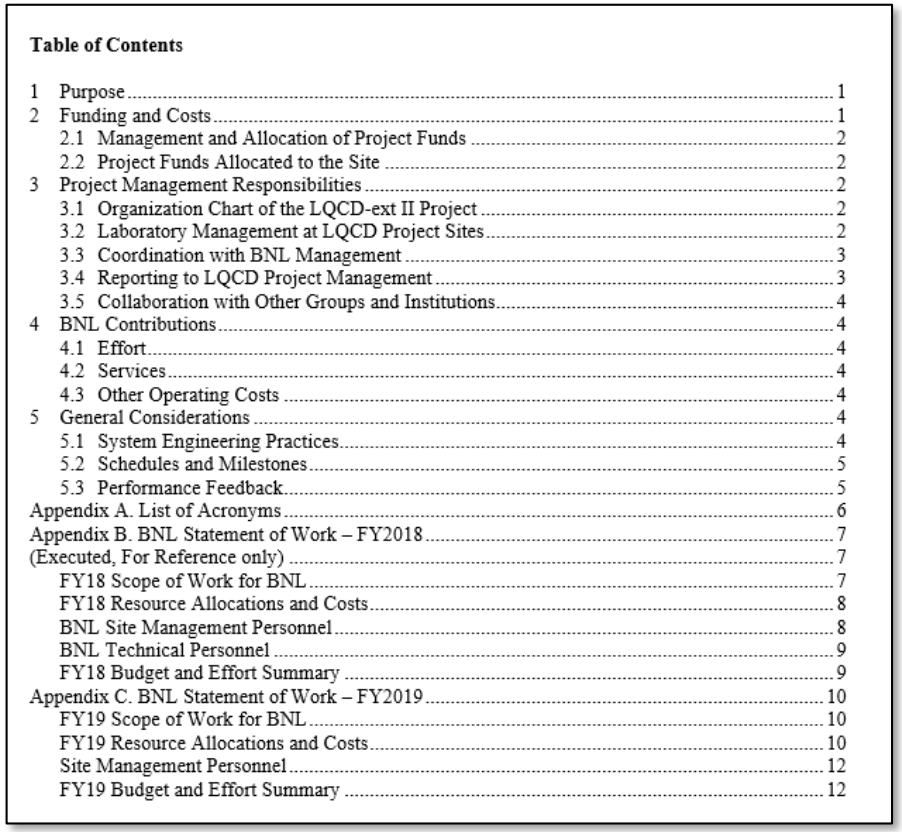


Table of Contents

FY19 Hardware Allocation Summary

FY19 Computing Resource Allocations

System	# Nodes	# Node-hrs	# Sky -core-hrs (millions)
BNL-IC	72	609,293	42.96
BNL-KNL	58	492,977	17.75
BNL-SL	58	492,977	17.75
FNAL-Pi0	289	2,467,135	23.27
FNAL-Pi0g	30	256,104	15.49
FNAL-LQ1*	69	237,967	7.61
Total	576	4,556,453	124.82

*LQ1 allocation duratoin was five months; all other allocation durations were for 12 months.

Other Contributions

- Besides providing personnel, computing cycles, and data storage facilities, each of the host laboratories provides additional services to the LQCD project:
  - Technical expertise (non-LQCD)
  - Executive consultation and advice
  - Networking services
  - Cyber security services
  - Web services
  - Administrative support

FY19 Hardware Activities

FY19 Hardware Acquisition Activities

- In July 2018, acquisition planning commenced on the FNAL FY19 institutional cluster procurement.

- A joint evaluation committee was formed to ensure the requirements of USQCD and Fermilab user groups were understood and properly considered in system design and specifications.
  - Committee charge co-authored by Bill Boroski (LQCD PM) and Liz Sexton-Kennedy (FNAL CIO) and issued on Jul 27, 2018.

    "Fermilab and the LQCD Computing Project are collaborating on the design, procurement, and installation at Fermilab of a high-performance computing cluster that will 1) meet the computing needs of LQCD and the Fermilab scientific community; and 2) be operated as an institutional cluster. The purpose of this committee is to understand user needs and existing computing resources and make a recommendation on the design and specifications for a new institutional compute cluster at Fermilab."

  - **The FY19 Acquisition Evaluation Committee**
    - *Robert Edwards (U), Chris Jones (F), Bob Mawhinney (U), James Osborn (U), Gabe Perdue (F), Amitoj Singh (Chair, representing both LQCD and FNAL)*

FY19 Hardware Acquisition Activities (2)

- Under the leadership of Amitoj Singh, the FY19 Joint Acquisition Evaluation Committee:
  - gathered requirements,
  - developed suitable code benchmarks,
  - considered existing resources and near/long-term demand,
  - evaluated viability of alternate computing architectures and production code availability,
  - provided a recommendation for the FNAL FY19 IC acquisition. (Report posted on review website)
- In parallel with evaluation committee activities and under the leadership of Amitoj Singh (FNAL), the FY19 LQCD-ext II alternatives analysis process was executed.
  - New hardware alternatives were defined and analyzed.
  - Activities, results and recommendation were recorded in the FY19 Alternatives Analysis (AA) document. (AA document posted on review website)
- After significant work, the FY19 Joint Acquisition Evaluation Committee recommended that we implement Alternative #1 in the FY19 Alternatives Analysis. (Sep 2018)

FY19 Acquisition Results

Implemented Solution

- A 90/10 (by budget) mixture of conventional and GPU-accelerated clusters released to production by March 29, 2019.

  Deploy and commission a conventional cluster of ~89 nodes and a GPU-accelerated cluster of ~2 hosts (4 GPUs per host, total 8 GPUs) capable of delivering respectively at least 39 TF and 11 effective TF, 50 TFlops total, with at least a memory capacity of ~12TB for a total M&S cost of $1.2M.

Acquisition Timeline

Activity	Date
RFP issued	Feb 22
RFP evaluations complete	Mar 29
PO awarded	Apr 12
First node received	Jun10
Installation complete	Jun 14
Released to production	Jul 1
Lustre storage installation complete	Jul 18 (est.)

Actual System Configuration

- 112 conventional Intel Xeon “Cascade Lake” nodes and 4 NVIDIA Tesla V100 nodes

FY19 Computing Resource Portfolio (as of 01-Jul-2019)

At BNL

- KNL Cluster (alloc: 58 of 72 nodes)
  - Intel Xeon Phi 7230 CPU (64 cores), 16 GB RAM on chip, 1.3 GHz
  - 2 x 512 GB SSD (with 512 MB internal buffer) for local storage
  - 192 GB DDR4 dual-rank RAM
  - Dual-rail (2x) Intel Omni-Path Host Fabric Interface Adapter 100 series
  - Intel TOR Omni-Path switches 1) dual-rail, non-blocking 2) 400 Gbps peak aggregate bi-directional bandwidth
- SL Cluster (58 of 64 nodes)
  - Two Intel Xeon "Skylake" Gold 6150 CPU (36 total cores), 25 MB Cache, 2.7 GHz
  - 1 x 4 TB SATA (6 Gbps) disk drive for local storage
  - 192 GB DDR4 dual-rank RAM
  - Infiniband EDR Host Fabric Interface Adapter VPI QSFP28
  - Mellanox non-blocking Infiniband EDR switches
- Institutional Cluster (72 of 144 nodes)
  - Dual-socket Broadwell CPUs
  - 2 NVIDIA K80 or P100 GPUs per node
  - 128 GB of memory per node
  - EDR Infiniband interconnect
- 400 TB GPFS disk storage
- 600 TB tape storage (in progress)

At FNAL

- Pi0 Cluster (alloc: 289 of 314 nodes)
  - Eight-core, dual-socket 2.6 GHz Intel Xeon (Ivy Bridge) nodes
  - 16 cores per node
  - 128 GB memory per node
- LQ1 Cluster (69 of 112 nodes)
  - Dual socket Intel Xeon "Cascade Lake" 6248 2.5GHz, 20-cores/socket
  - 192GB DDR4-2933 per node
  - ~1TB local disk (/scratch)
  - Intel Omni-Path EDR (100 Gbps)
  - ~1PB Lustre storage
- Pi0g Cluster (30 of 32 nodes)
  - Eight-core, dual-socket 2.6 GHz Intel Xeon (Ivy Bridge) nodes
  - 128 GB memory per node
  - 4 GPUs NVIDIA K40m (Kepler Tesla) per node (120 total GPUs available)
  - GPU memory (ECC on) 11.5 GB/GPU
- 600 TB Lustre Disk Storage
- 1 PB Tape Storage

FY19 Performance Results to Date

LQCD-ext II Progress: FY19 Year-to-Date

- FY19 (thru May 2019)
  - BNL brought online 600 TB of new tape storage and expanded the size of their IC machine by 54 nodes, from 162 to 216 nodes
    - Current configuration: 50% K80s, 50% P100s.
  - FNAL brought online a new cluster (LQ1) and additional storage (1PB Lustre) on July 1.

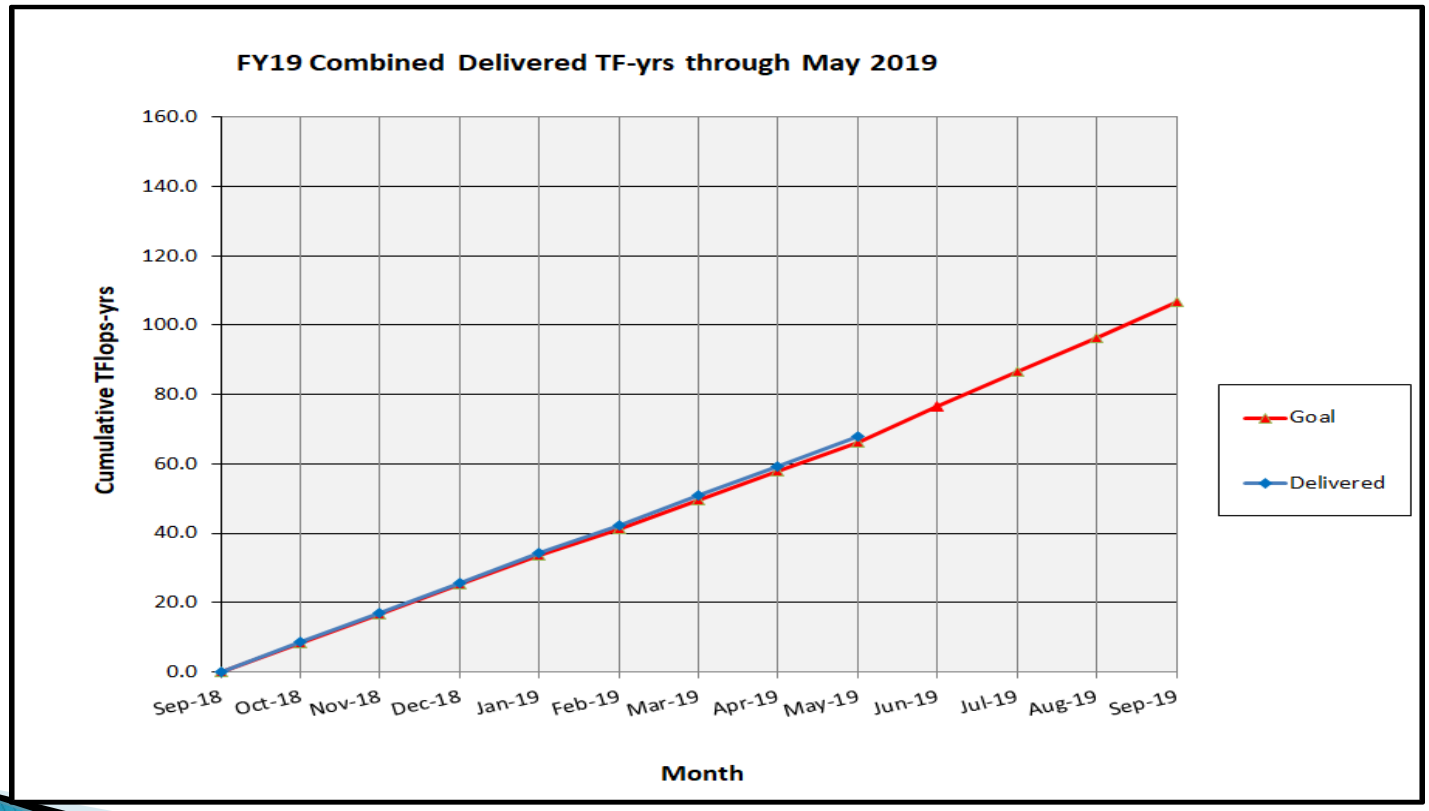
Delivered Computing (TF-yrs)						
	FY19 (Oct '18 thru May '19)			Cumulative (Oct '14 thru May '19)		
	Goal	Actual	% of Goal	Goal	Actual	% of Goal
Conventional Resources	29.5	30.5	103%	451.2	474.1	105%
Accelerated Resources	36.8	37.4	102%	386.9	403.8	104%

1) Conventional resources operational in FY19: BNL: KNL, SL; FNAL: Pi0, LQ1
DNR Mode @ FNAL: (Bc, Ds)

2) Accelerated resources operational in FY19: BNL: IC; FNAL: Pi0g
DNR Mode (Dsg

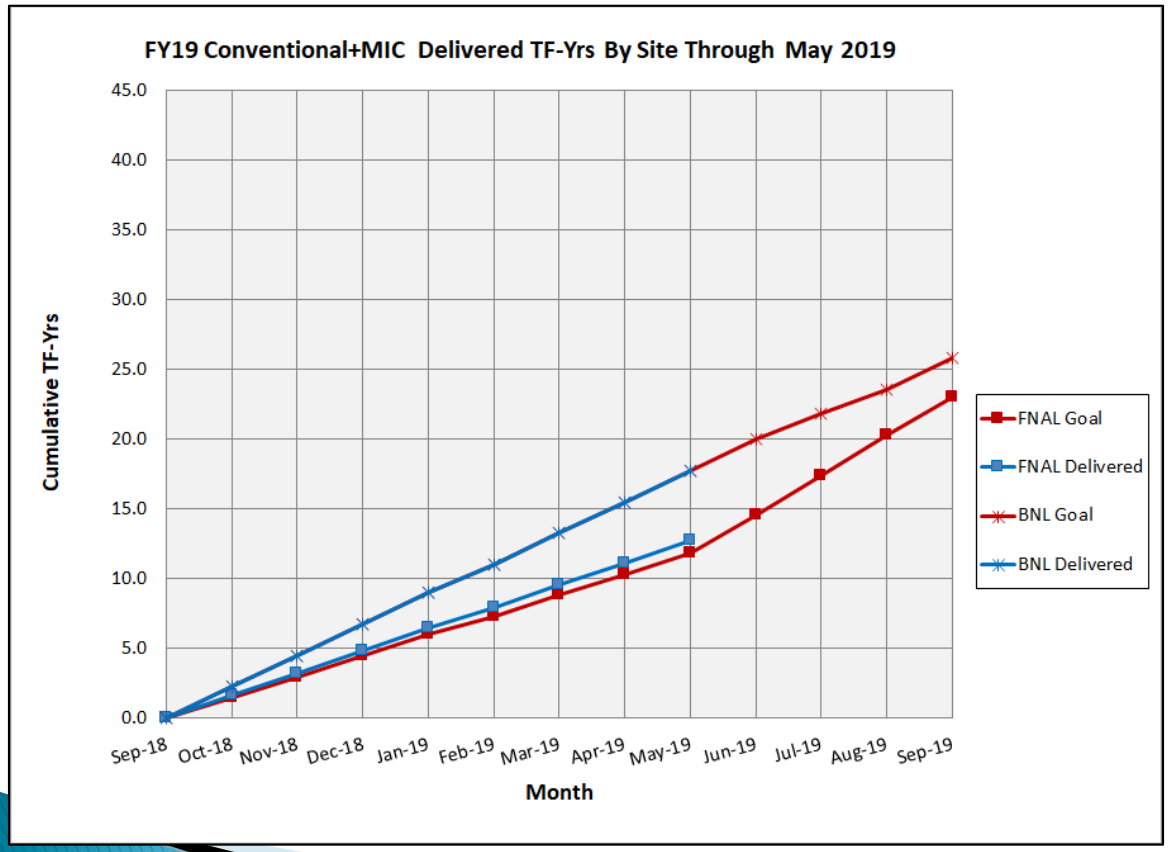
FY19 Milestone Performance – (TFlops-yrs delivered) Combined Resource Performance

- Computing delivered in FY19 from conventional and GPU compute hardware is shown.
- Performance goal through May 2019: 66.3 TFlops-yrs.
- The project achieved 67.9 TFlops-yrs (102% of goal).
- 100% availability of allocated node counts on all systems.



FY19 Milestone Performance – (TFlops-yrs delivered) Conventional + MIC Hardware

- Computing delivered in FY19 from conventional compute hardware is shown.
- Through May 2019: Goal=29.5 TFlops-yrs; Actual = 30.5 TFlops-yrs (103% of goal)



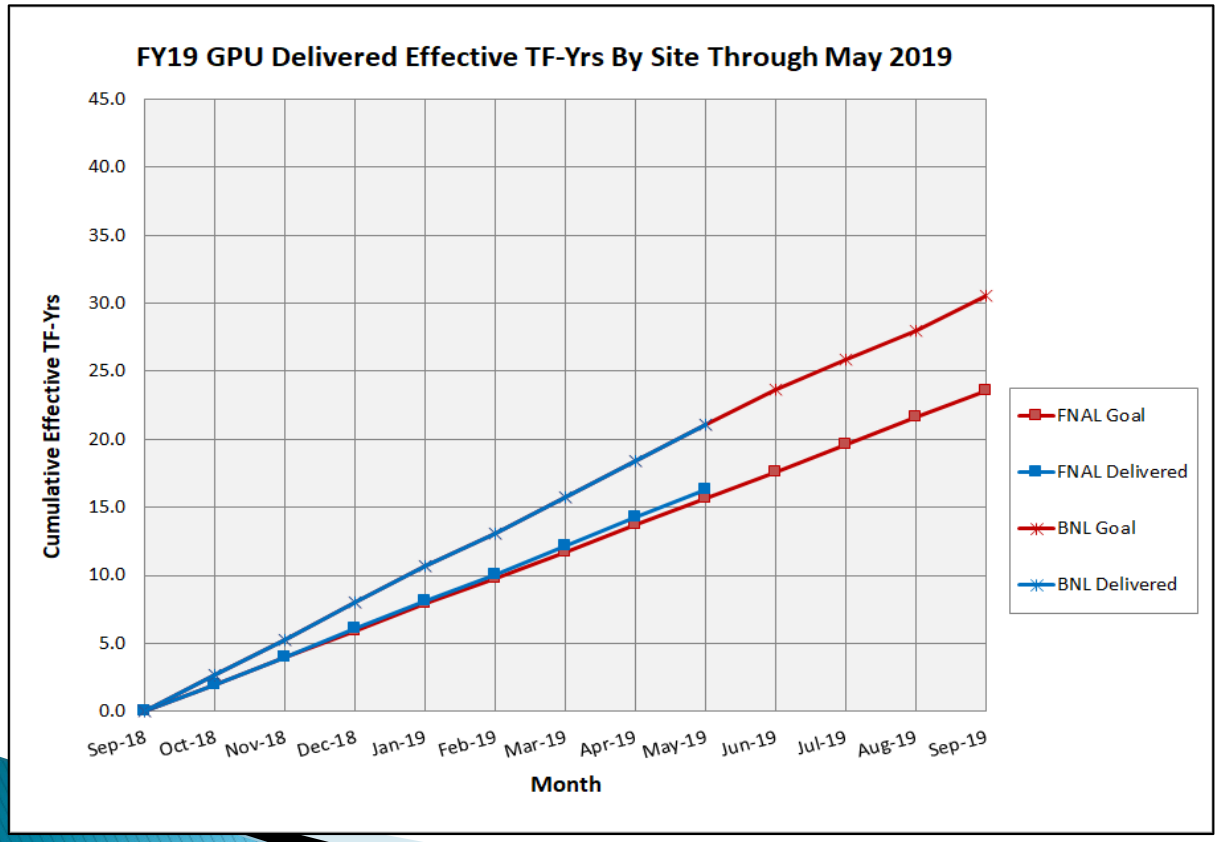
Conventional computing resources:

- FNAL: Pi0
- BNL: KNL
- BNL: SL

The inflection point in the Uptime Goal curve corresponds to the deployment of the new Institutional Cluster at FNAL Planned deployment was in May; actual production release was July 1.

FY19 Milestone Performance – (TFlops-yrs delivered) Accelerated Hardware

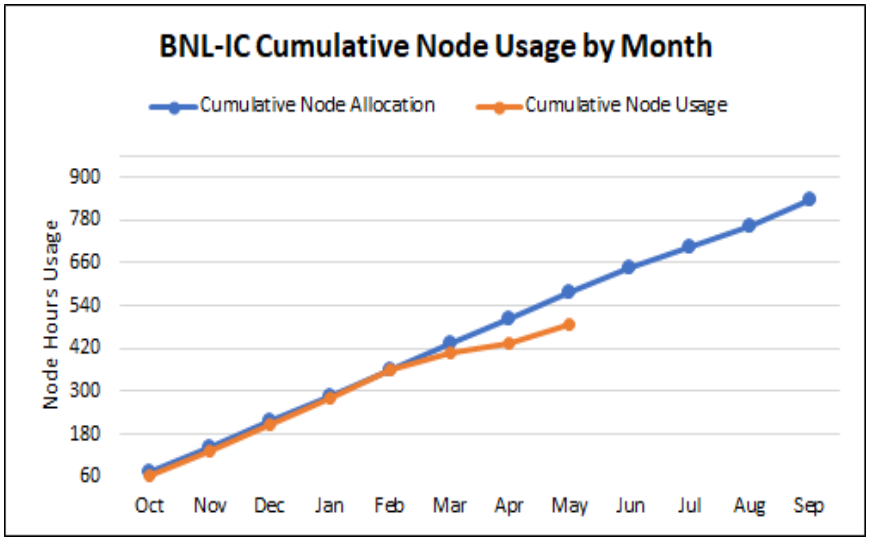
- Computing delivered in FY19 from accelerated compute hardware is shown.
- Conversion from GPU-hrs to effective TF-yrs is 140 GF/GPU, based on allocation-weighted performance of GPU projects running from July 2012 to December 2012.
- Through May 2019: Goal=36.8 TFlops-yrs; Actual = 37.4 TFlops-yrs (102% of goal)

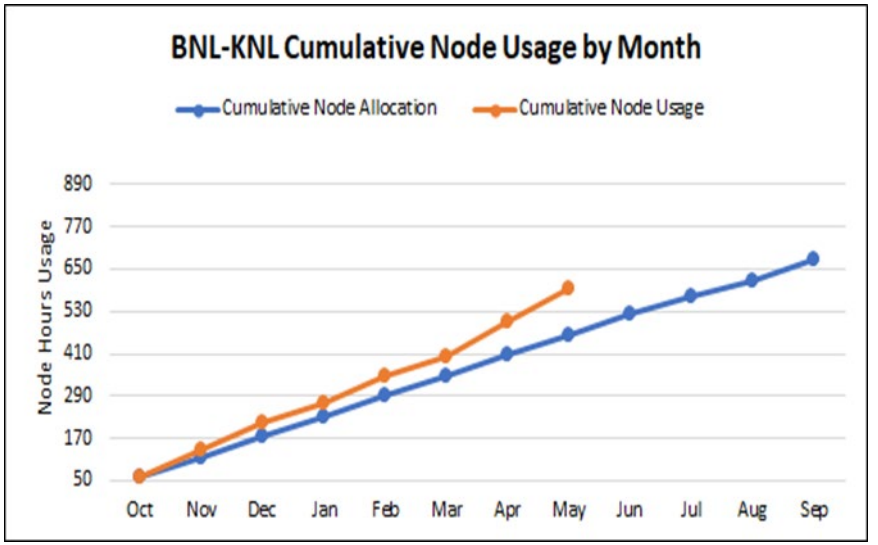


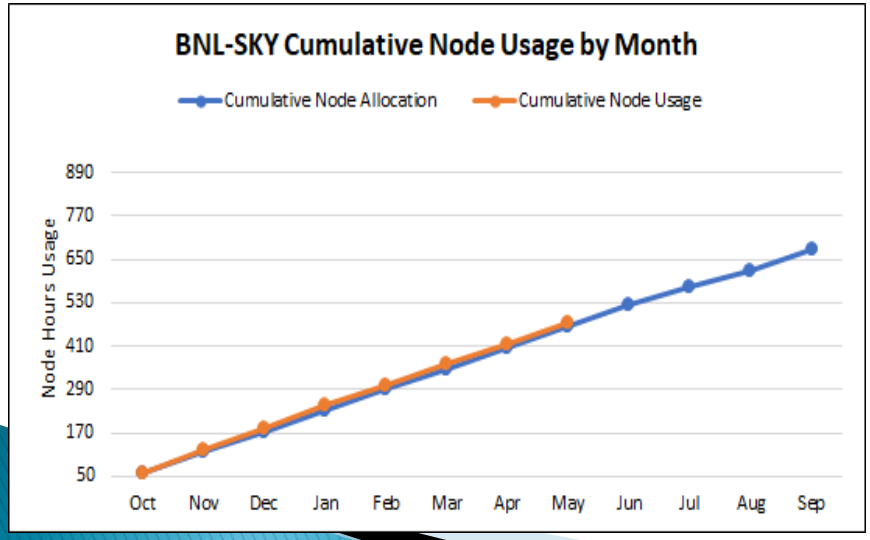
Accelerated computing resources:

- FNAL: Pi0g
- BNL: IC

Utilization of our FY19 Allocations – BNL



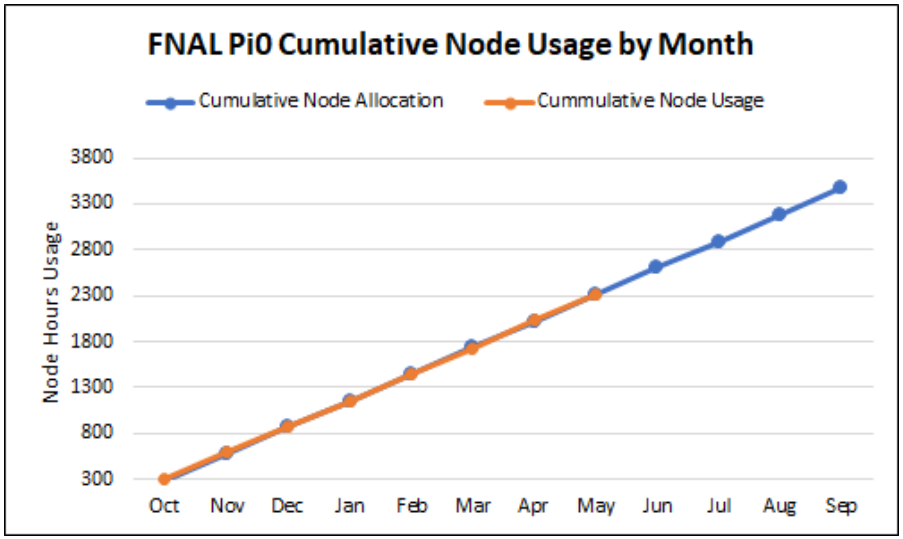


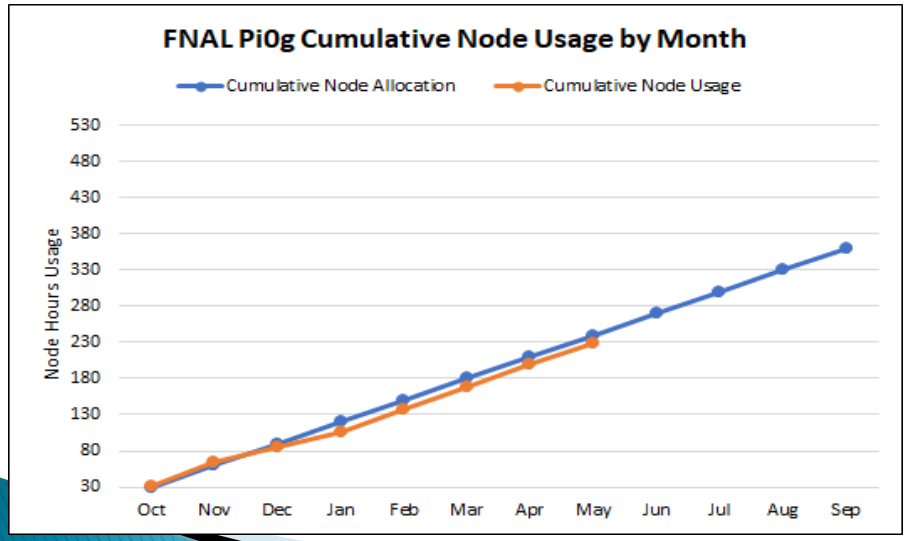


FY19: Oct thru May:

System	Node Allocation	% of Cumulative Allocation Used
BNL-IC	72	84%
BNL-KNL	58	115%
BNL-SKY	58	103%

Utilization of our FY19 Allocations – FNAL





FY19: Oct thru May:

System	Node Allocation	% of Cumulative Allocation Used
FNAL-Pi0	289	100%
FNAL-Pi0g	30	95%

FY19 May Utilization – FNAL

Planned FNAL power outage

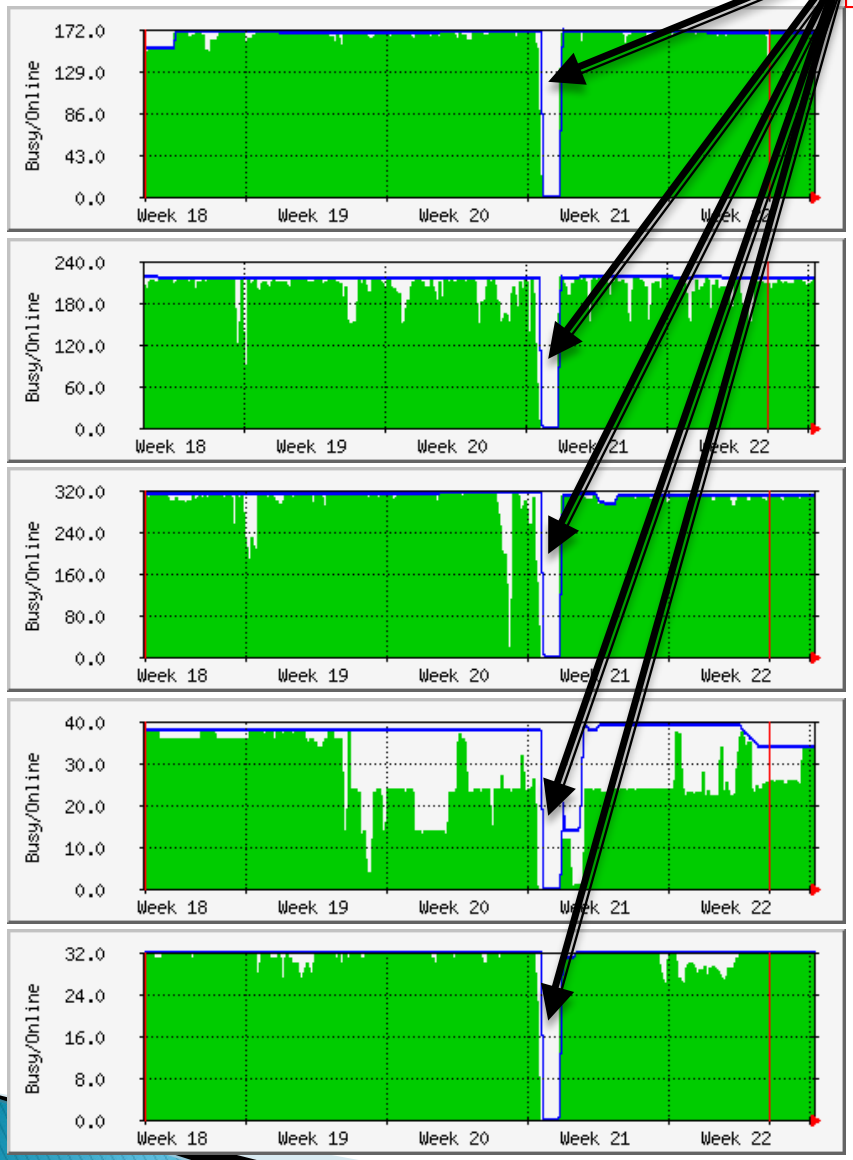
Ds

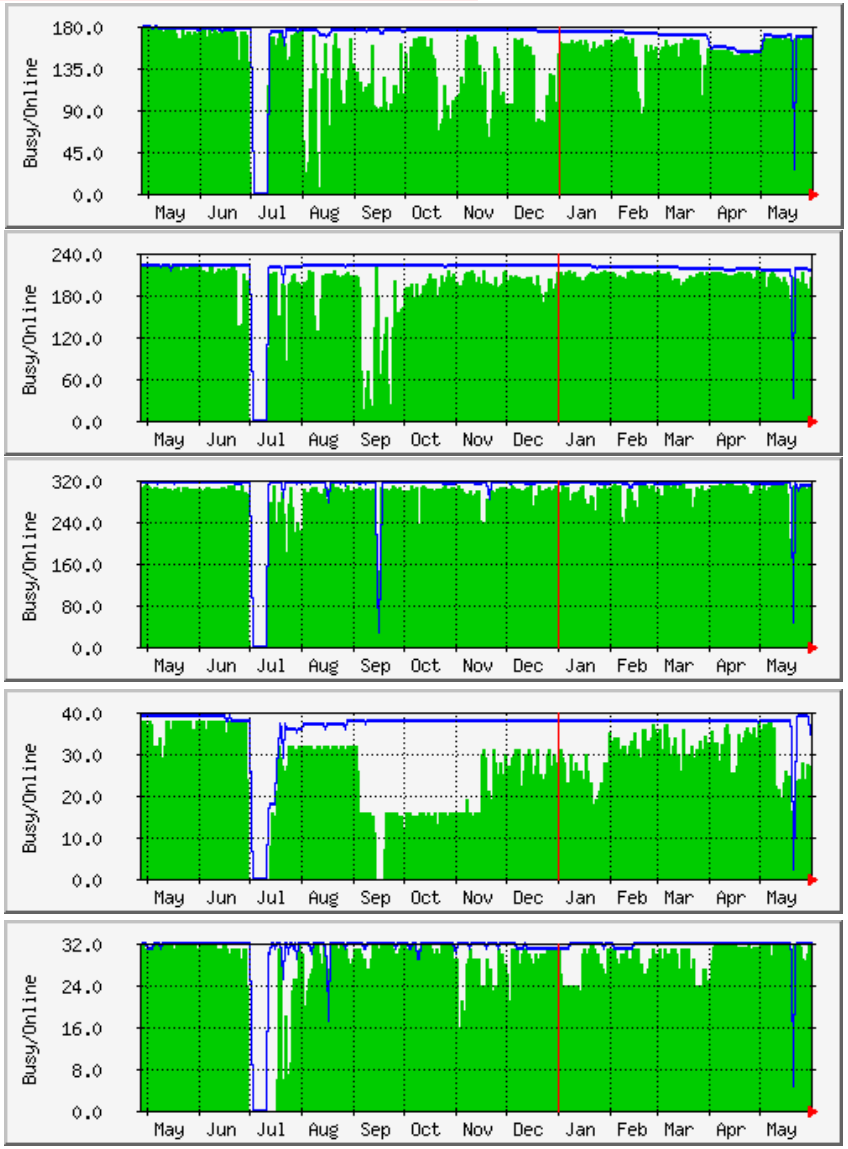
Bc

Pi0

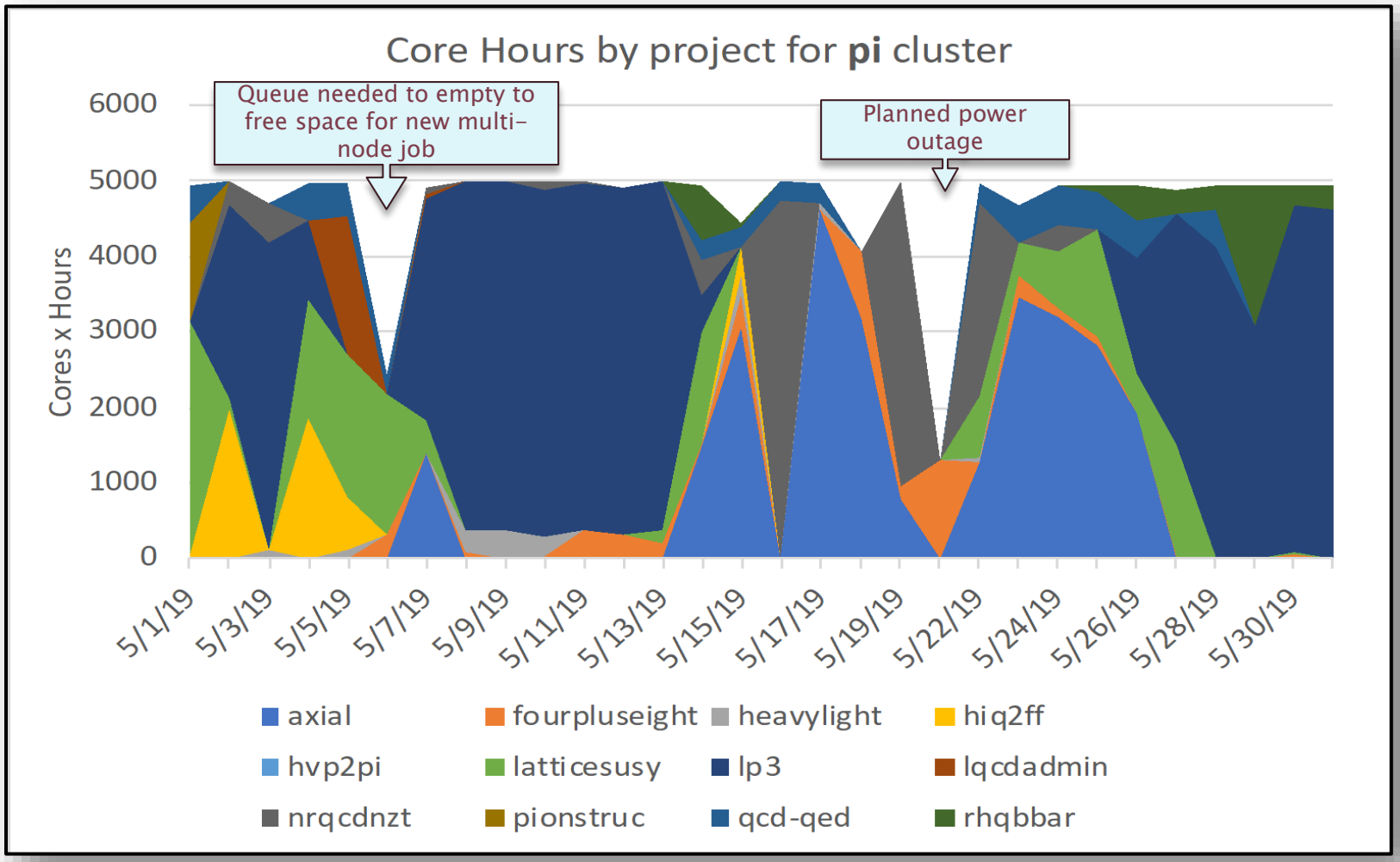
Dsg

Pi0g





FY19 May Utilization – FNAL Pi Cluster



Project Financials: FY19 Costs through May 2019

Cost Performance Summ* *FY completed:* *67%* *Period Begin/End: 10/1/2018- 9/30/2019

Operating and Equipment Funds

Lab	FY19 Budget	FY19 YTD Costs	FY19 YTD Open Commitments	FY19 YTD Balance	% Spent & Committed
BNL	$1,654,325	$1,106,758	$0	$547,567	67%
FNAL1	$2,122,530	$1,640,638	$58,277	$423,615	80%
Sub-total	$3,776,855	$2,747,396	$58,277	$971,182	74%
Mgmt Reserve	$0	$0	$0	$0	0%
Total	$3,776,855	$2,747,396	$58,277	$971,182	74%

Notes:
1) Included total carryover of $425,110 (OP) and $760,000 (EQ).

- Spending is in line with the expectations
- FNAL YTD Open Commitment dollars include: 3 storage servers, 2 rackmount servers, and 3 cluster servers for the new Fermilab Institutional Cluster (LQ1)

FY19 Performance Summary

- Operations across the board continue to run very smoothly.
- We are on track to exceed delivered computing goals for the year.
  - Through May 2019, we have delivered 67.9 TF-yrs against our target of 66.4 TF-yrs. (102%)
- Usage of the LQCD facilities by the USQCD community is heavy. We are on track to consume FY19 allocations.
- We brought online new compute and storage capacity as planned.
  - Deployment of the new FY19 institutional cluster was late due to procurement process and delay in overseas shipment for equipment. Production release occurred on July 1.
  - Due to favorable pricing, the size of the new cluster is larger than expected (112 conventional nodes vs. 89 nodes, plus 4 GPU nodes vs. 2).
- MOUs are in place with the host laboratories and all parties are meeting or exceeding deliverables and expectations.
- We have very strong working partnerships with our host laboratories and USQCD stakeholders.

Questions?