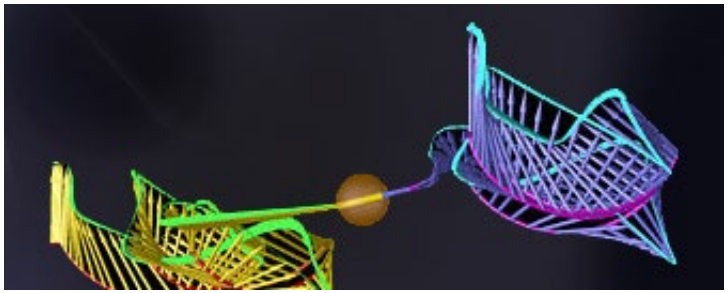


## LQCD Research Program Extension (LQCD-ext III)

# Institutional Cluster Computing and Operations Model

Bill Boroski
LQCD-ext Contractor Project Manager
boroski@fnal.gov

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

# Outline

- Scope, Operating & Acquisition Models, and Budget
- Management and Oversight
  - Organizational Structure
  - Work Planning
  - Risk Management
  - Cyber Security
  - Quality Assurance
- Performance Management
  - Key Performance Indicators
  - Cost and Schedule Performance

# Program Scope

- Support the extension of the LQCD research program by procuring and coordinating mid-scale computing resources required by LQCD for fiscal years 2020-2024.
  - Continue to acquire compute cycles provided by institutional computing hardware at existing facilities located at BNL and FNAL, using a computing services model.
  - Partner with BNL and FNAL on the acquisition and deployment of new institutional computing hardware that meets ongoing and future needs of LQCD and other science programs.

- Out of scope: software development, scientific software support

Operations Strategy

- Utilize mid-scale institutional computing resources to meet the computational needs of the LQCD research program.
- Partner with the host laboratories to ensure that computing resources are allocated as required and delivered as planned; and fully utilized by the LQCD user group.

Acquisition Strategy

- Partner with the host laboratories to ensure that the design, architecture, and performance of new mid-scale institutional computing resources meet the computational requirements of the research program.

# Operations Model

- The LQCD Program Office will manage and oversee all program activities. The Contractor Program Manager will be the primary point of contact with the DOE Federal Director.
- The LQCD Program Office will maintain frequent communication with the USQCD Executive Committee (EC) and Scientific Program Committee (SPC) leadership to ensure user needs are well understood and appropriately met.
- Computing and storage needs will be determined by the USQCD Scientific Program Committee through the annual resource allocation process.
- Memoranda of Understanding (MOUs) will be established with each host laboratory. Statements of Work (SOWs) will be created to document annual resource commitments and related obligations.
- LQCD-ext III site managers will manage and oversee site activities, including user account creation, resource allocations, performance monitoring and tracking, budget planning, and cost tracking.

# Steady-state Operations & Maintenance

- User allocations are determined annually by the Scientific Program Committee and provided to each site manager for implementation
- Site Managers are responsible for day-to-day operations of their respective sites. Responsibilities include:
  - Establishing systems to track system performance and usage;
    - Utilization
    - Uptime and Delivered TFlops-yrs
    - Job failure rates and time lost to failed jobs
    - Progress against allocations
  - Reporting progress against goals;
  - Identifying issues and concerns to the CPM;
  - Monitoring the acquisition and deployment of new systems at their institution.
- Services provided by the site staff to users
  - New account requests
  - Storage management (quotas, critical area backup/restores, etc.)
  - Helpdesk
  - Web pages providing user documentation, system status, etc.
  - Special requests (e.g., high priority queues)

# Acquisition Model

- The LQCD-ext III Program Office will maintain a 5-year hardware portfolio roadmap
  - Define current and planned mid-scale computing assets available to the LQCD research program from the host laboratories.
  - Production lifecycle for suitable clusters is typically 5 years.
- The Program Office and USQCD leadership will meet with laboratory computing leadership on an annual basis for planning purposes
  - Review LQCD computational needs, vendor and LCF roadmaps, institutional hardware roadmaps, future acquisition plans, etc.
- For laboratory acquisitions applicable to the LQCD research program, a joint evaluation committee will be formed to evaluate options
  - The joint committee will consist of subject matter experts (SMEs) representing the needs of LQCD and laboratory user groups.
  - Gather requirements, identify potential solutions, formulate code benchmarks, and recommend a preferred solution.
  - Present results in a written report.

# Acquisition Model (2)

- The Program Manager will approve recommended solutions with concurrence from the Executive Committee
  - Notification is also provided to the DOE Federal Director
- The host laboratory will procure, install, and deploy new systems
  - Deployment schedule will be communicated to LQCD Program Office and tracked to completion
  - LQCD users will participate in user acceptance testing via benchmarks and friendly user mode
- Lessons learned will be gathered by the Program Office and shared with the host laboratory in the spirit of continuous improvement.

# Institutional Clusters

- Refers to a system of interconnected mostly-homogenous compute nodes of a common architecture, with associated network switches, storage, etc., and managed using a single job scheduler (e.g., SLURM).

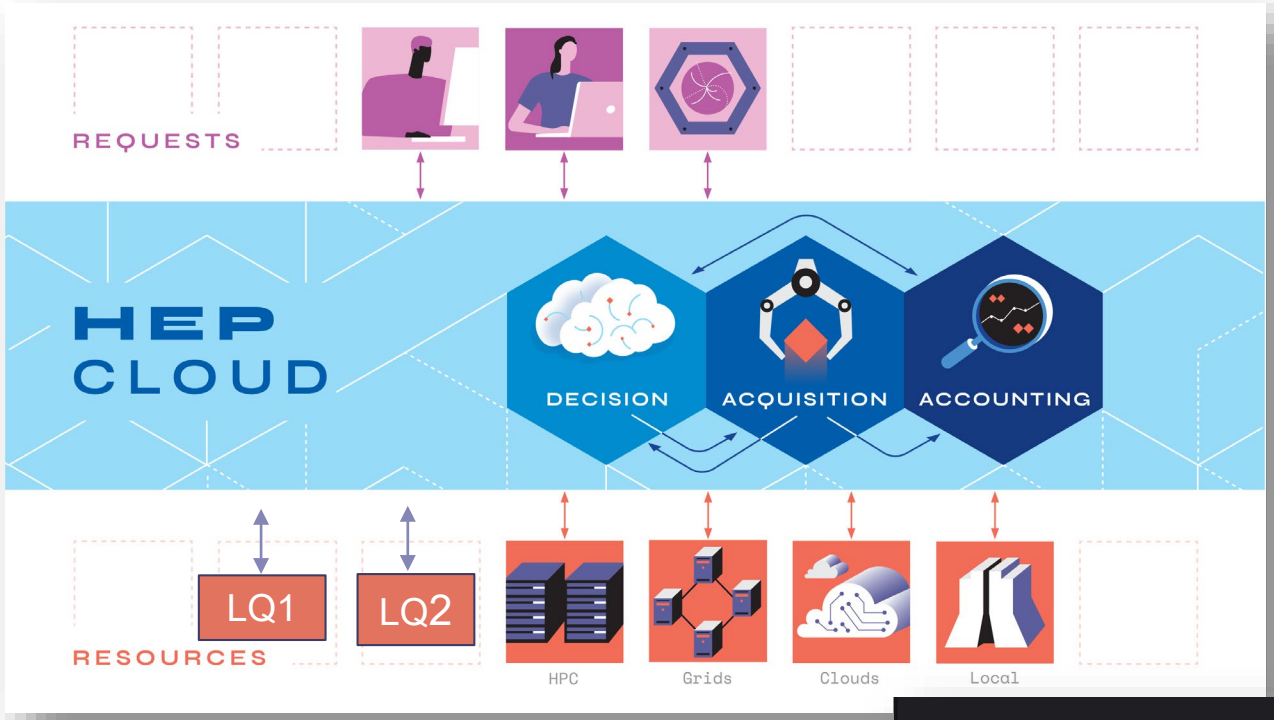
- Designed to serve multiple user groups
  - For LQCD-ext II, adoption has required a move away from dedicated systems specifically designed and optimized for LQCD codes and needs (e.g., low-latency interconnects)
  - LQCD hardware portfolio currently consists of a mix of ICs and dedicated hardware

- Locally managed following host laboratory policies and procedures (i.e., security, ES&H)

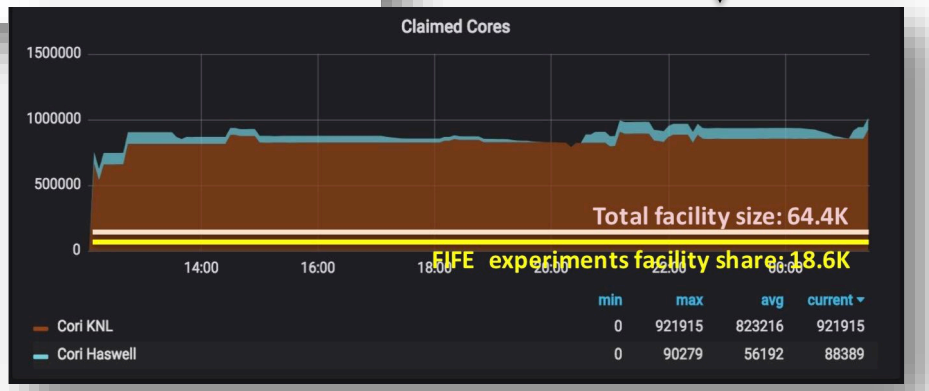
# Institutional Clusters (2)

- Designed, procured, installed and maintained locally by common support groups
  - Deeper bench for knowledge retention, staff coverage, cross-training, etc.
  - Slightly less direct personal attention.
  - Centralized service desk ticketing systems and response procedures following best-practice service management processes.
- Service models based on cost per unit delivered
  - Computing: cost per node-hour ($/node-hr.).
  - Data storage: cost per storage unit ($/TB [disk] or $/PB [tape])
  - Pay for what you need
- Advantages
  - Potential lower overall operating costs due to economies of scale and scope
  - Ability to adjust architecture allocations as needs change
  - Large institutional clusters have the capacity to balance ebb and flow of different user group demands
- Aligned with the strategic direction for scientific computing at host labs

# Future Fermilab Institutional Cluster Model



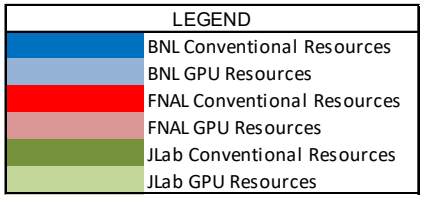

NOvA using HEPCloud to claim ~1M cores at NERSC to perform a large-scale analysis over a short timeframe.




Resource demand spikes that cannot be met by on-premise resources can be satisfied by routing jobs to commercial clouds, OSG resources or LCF's.

# Existing Resources & Future Plans

- LQCD-ext III will use existing IC resources at BNL and FNAL in FY20, and new resources as existing systems are expanded, or new systems are brought on-line.
  - FNAL plans to expand its IC beginning in FY20
  - BNL does not plan to expand current IC. BNL is outfitting a new data center, with a new GPU-based IC put in place in FY22. BNL also purchases small prototype systems for CSI-related R&D projects.
- LQCD resources at JLab are available to USQCD and included in the annual allocation process. JLab hardware architectures are considered when planning new acquisitions to meet LQCD user needs. They are part of the overall hardware portfolio.




**Available Cluster Systems**

| System | FY20 | FY21 | FY22 | FY23 | FY24 |
|---|---|---|---|---|---|
| Conventional Resources | | | | | |
| BNL-KNL | | | | | |
| BNL-SL | | | | | |
| FNAL-LQ1 | | | | | |
| FNAL-new CPU | | | | | |
| BNL-new CPU | | | | | |
| JLab-xx (KNL) | | | | | |
| GPU Resources | | | | | |
| BNL-IC | | | | | |
| FNAL-new GPU | | | | | |
| BNL-new GPU | | | | | |
| JLab-18g (GPU) | | | | | |
| **Total Computing (TF-yrs)** | **64.88** | **76.17** | **85.09** | **113.55** | **120.70** |

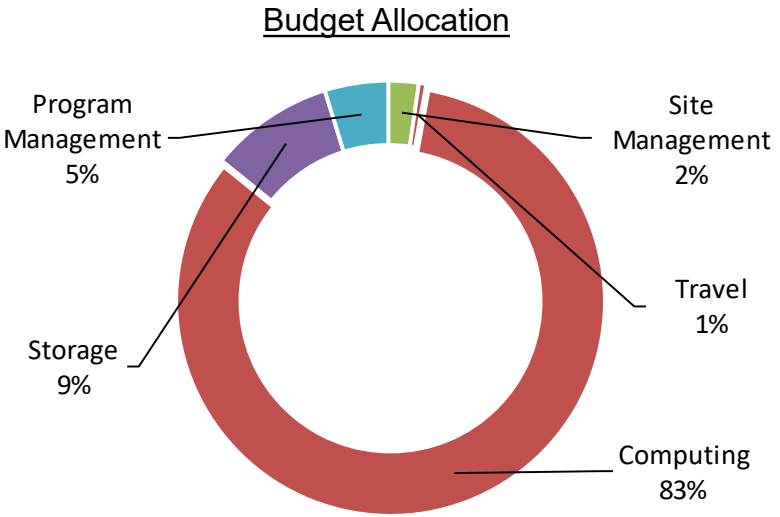
Notes:
1) Total Computing only includes LQCD-ext III resources (BNL & FNAL)
2) JLab systems are operated under a dedicated hardware model.

# Program Budget

- Obligation budget = $10.82 M
  - *Based on guidance from DOE Office of High Energy Physics (HEP)*
- Period of performance: FY20 through FY24

| | FY20 | FY21 | FY22 | FY23 | FY24 | Total |
|---|---|---|---|---|---|---|
| Budget ($M) | 2.030 | 2.095 | 2.165 | 2.230 | 2.300 | 10.820 |

- Program funding covers:
  - Procurement of compute cycles from existing and future Institutional Clusters operated at BNL and FNAL.
  - Procurement of disk and tape storage capacity to support science program
  - Program management
  - Site management
  - Travel
- Not in scope
  - Software development
  - Scientific software support



# Budget Allocation by Expenditure Type

Planning Budget = $11.6M

Includes projected unspent funds from LQCD-ext II due in part to late release of new FNAL institutional cluster (LQ1)

| Expenditure Type | FY20 | FY21 | FY22 | FY23 | FY24 | Total |
|---|---|---|---|---|---|---|
| Site Management | 50 | 52 | 53 | 55 | 56 | 265 |
| Travel | 12 | 12 | 13 | 13 | 14 | 64 |
| Computing | 1,889 | 2,223 | 1,776 | 1,833 | 1,898 | 9,620 |
| Storage | 217 | 217 | 218 | 219 | 220 | 1,090 |
| Program Management | 125 | 103 | 106 | 109 | 113 | 556 |
| Management Reserve | | | | | | |
| Total | 2,293 | 2,608 | 2,166 | 2,229 | 2,300 | 11,596 |

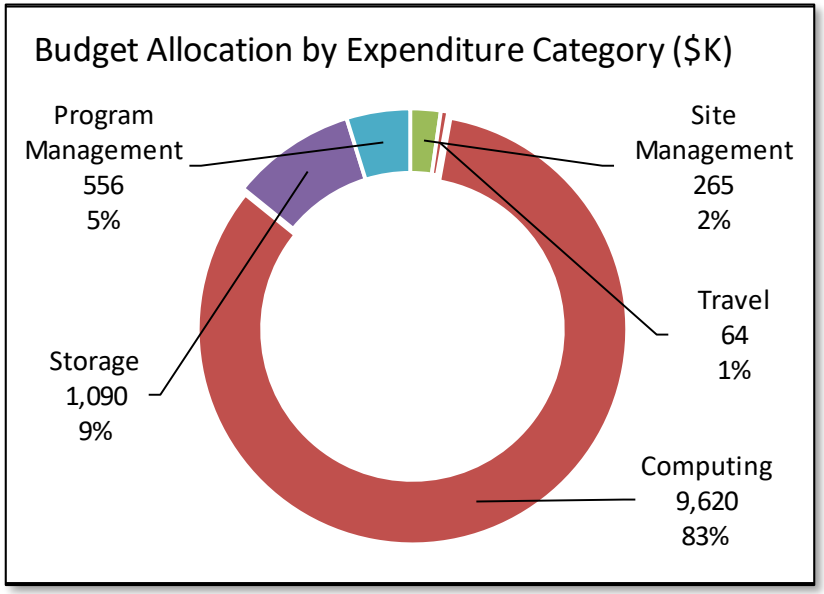
Site Mgmt: Primary liaison with lab, account mgmt, monitoring & reporting, etc.

Travel: All-hands Mtg; Annual DOE reviews

Computing: Procurement of node-hrs from IC facilities at BNL and FNAL.

Storage: Procurement of disk and tape storage usage at BNL and FNAL. Also covers tape service fees.

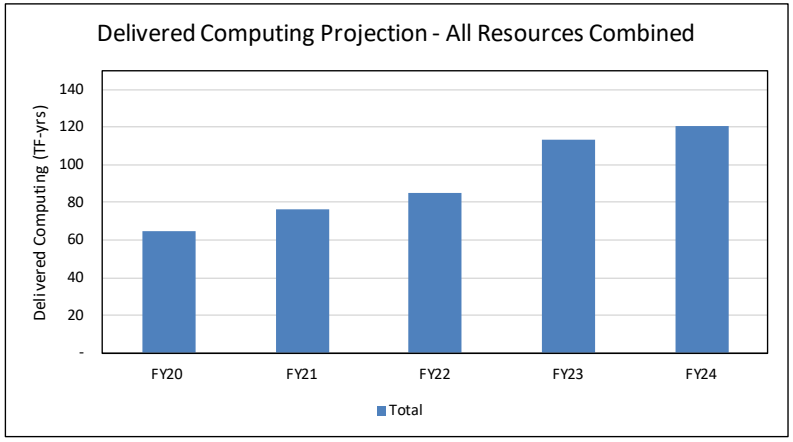
Prog Mgmt: Planning, mgmt, and oversight for computing program. Primary liaison with DOE and host laboratories.

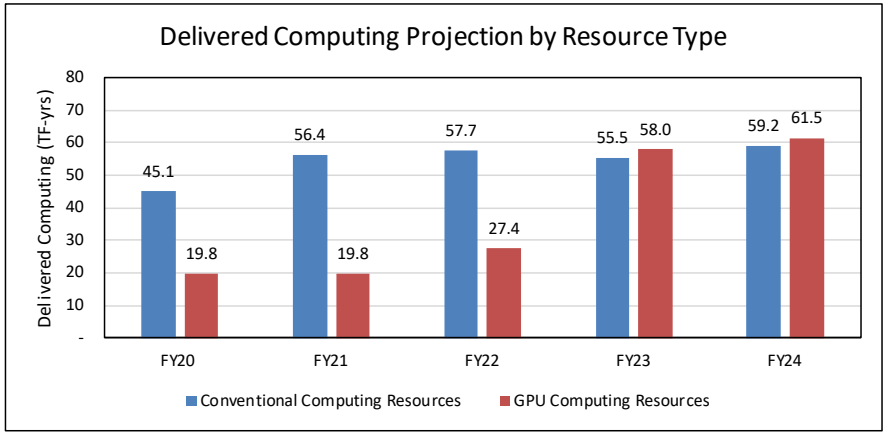


# Delivered Computing Projection

- Deployment and cumulative performance milestones defined for each year:
  - "Delivered TFlops–yrs"
    - Available capacity expressed as average of DWF and highly improved staggered quark (HISQ) algorithms.
    - 1 year = 8760 hours

| Estimated Computing Delivered (TF-yrs) | | | | | | |
|---|---|---|---|---|---|---|
| **Category** | **FY20** | **FY21** | **FY22** | **FY23** | **FY24** | **Total** |
| Conventional Computing Resources | 45.08 | 56.37 | 57.72 | 55.51 | 59.21 | 273.88 |
| GPU Computing Resources | 19.80 | 19.80 | 27.38 | 58.04 | 61.49 | 186.51 |
| **Total** | **64.88** | **76.17** | **85.09** | **113.55** | **120.70** | **460.39** |

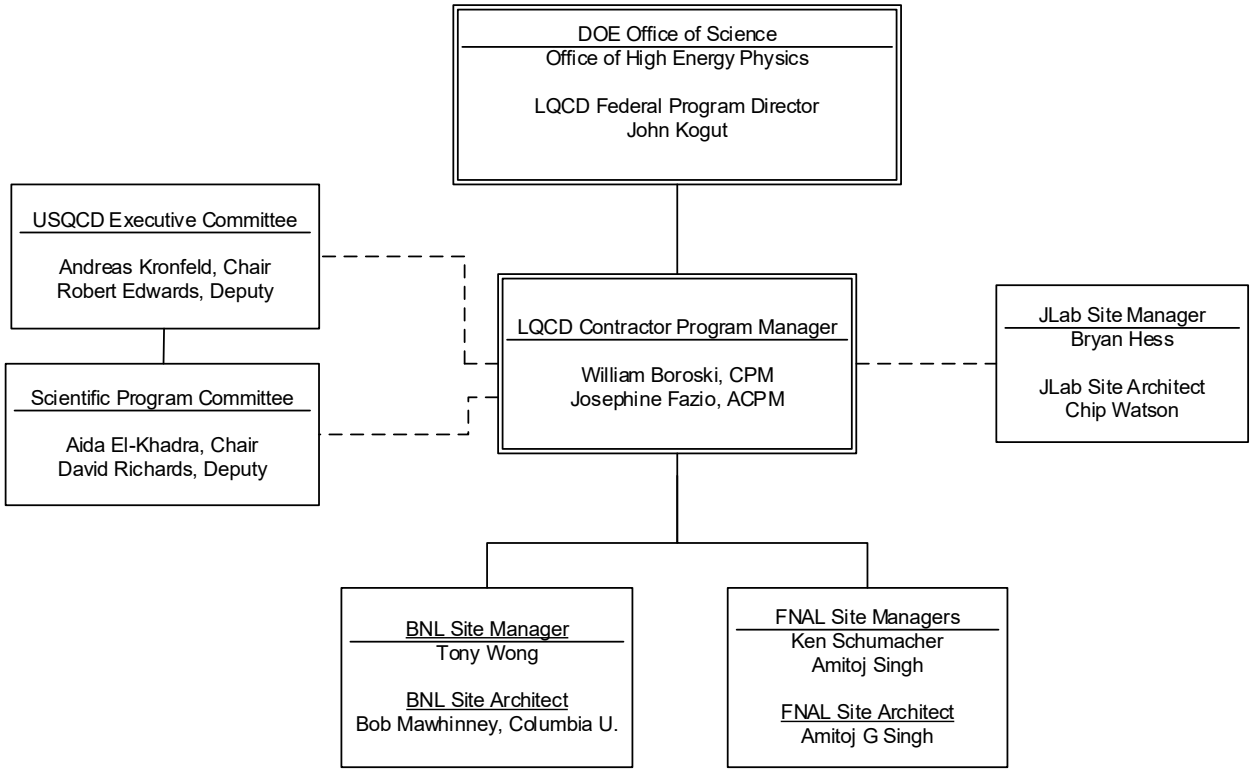



All deployment figures assume that 70% of the annual hardware budget is used to purchase conventional hardware cycles and 30% to purchase accelerated hardware cycles (e.g., GPUs).

# MANAGEMENT & OVERSIGHT

# LQCD-ext III Management Organization

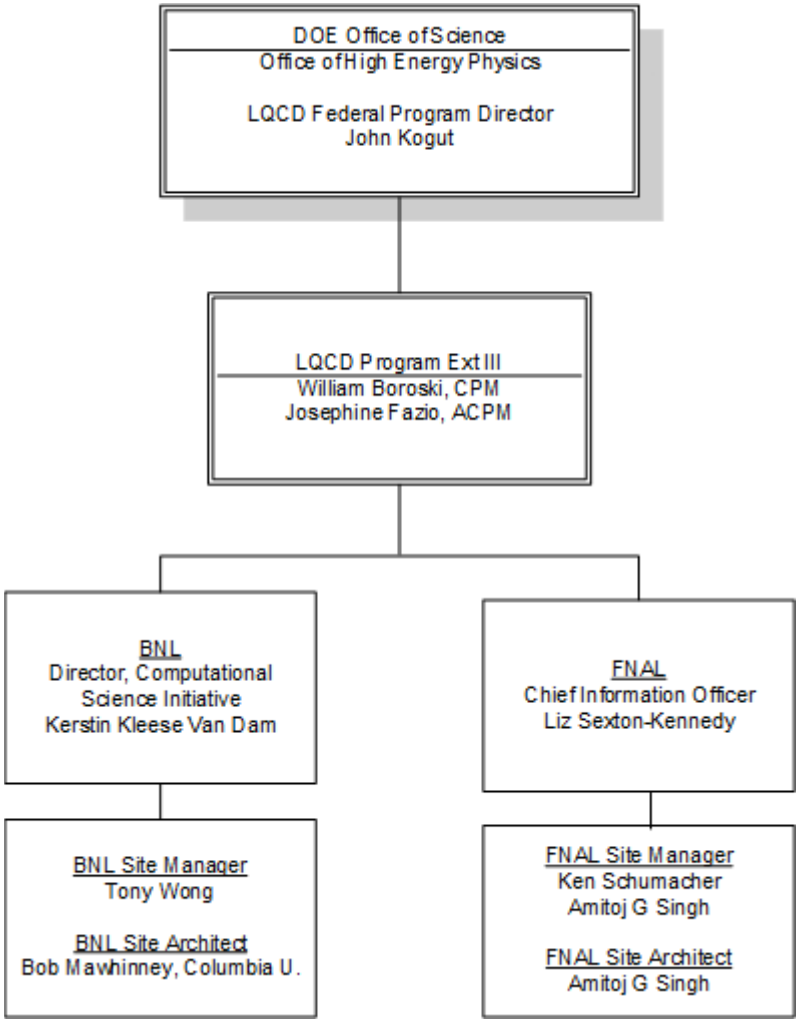



The management and oversight structure is the same as that used for the LQCD-ext II project. A similar structure was also used successfully to manage the LQCD-ext project.

The Federal Project Director (*Kogut*) and Contractor Project Manager (*Boroski)* are certified “Level 1 Qualified IT Project Managers”, in accordance with the DOE OCIO Project Management Qualification Requirements.

Roles and responsibilities for all key positions are defined in the Program Execution Plan

# Interaction with Host Laboratory Management



# Integrated Program Team (IPT)

- The Integrated Program Team (IPT) consists of stakeholder representatives and key personnel.
- Membership reflects and represents the interests of stakeholders, program, and user community.

- IPT:
  - John Kogut, Federal Project Director (OHEP)

  - Bill Boroski, Contractor Project Manager
  - Josephine Fazio, Associate Contractor Project Manager
  - Tony Wong, BNL LQCD Site Manager
  - Bob Mawhinney, BNL Site Architect
  - Ken Schumacher, FNAL LQCD Site Manager
  - Amitoj Singh, FNAL LQCD Site Architect

  - Andreas Kronfeld, USQCD Executive Committee Chair
  - Robert Edwards, USQCD Executive Committee Deputy
  - Aida El-Khadra, USQCD Scientific Program Committee Chair
  - David Richards, USQCD Scientific Program Committee Deputy

# Communications and Reporting

- Bi-weekly Site Managers Meeting
  - Participants: CPM, ACPM, Site Managers, USQCD Executive Committee (EC) Chair
  - Address site-specific issues or concerns and discuss procurement plans/activities
  - Exchange of other relevant information between technical, mgmt, and scientific staff

- Monthly DOE Program Office Meeting
  - Participants: Federal Project Director (OHEP), CPM, ACPM, EC Chair, EC Deputy
  - Report on progress against performance goals (TFlops-yrs delivered, cost, procurement activities, etc.)
  - General exchange of information related to program planning and performance

- Annual Progress Review
  - Review of scientific, technical, and program management performance by external review committee.
  - Organized by the Federal Project Director

- Annual USQCD All-Hands Meeting
  - Participants: CPM, ACPM, Site Managers, USQCD Executive Committee, Scientific Program Committee, USQCD Community
  - Exchange of information between the project team and collaboration regarding computing facility performance and plans, operational topics, scientific program needs, etc.

- Informal communications between Federal Project Director and Contractor Program Manager as necessary.

# Work Planning and Organization

- Program Execution Plan (PEP)
  - Controlled document defining scope, requirements, management roles and responsibilities, cost and schedule, change control, performance goals and metrics, etc.
- Work Breakdown Structure (WBS)
  - MS Project used to identify tasks, develop schedules, and track progress against milestones
  - Work broken down into two primary areas:
    - Steady-state operations and maintenance
    - Procurement and deployment of new systems
- Other important documents
  - Risk Management Plan, Quality Assurance Plan, Alternatives Analysis, Annual Acquisition Plans, Alternatives Analysis, C&A Documentation
- MOUs with host institutions

# Risk Management

- Risk assessment and analysis has been performed and is documented in the LQCD-ext III Risk Management Plan
  - Risks are considered and categorized into one of five major risk areas: Technology, Cost, Schedule, Security, or Service
  - During project execution, risk management is performed on an ongoing and continuous manner in accordance with DOE requirements (e.g., DOE O413.3B)
- Identified risks are captured and documented in a Risk Register that is maintained by the Associate Program Manager.
  - Each risk is assigned a risk rating based on probability of occurrence and severity of impact.

Table 1. Risk Probability and Impact Values

| Probability | Value | Impact | Value |
|---|---|---|---|
| High | 0.75 | Severe | 0.9 |
| Medium | 0.50 | Moderate | 0.5 |
| Low | 0.25 | Low | 0.1 |

Table 2. Risk Rating Matrix

| **Prob \ Impact** | Severe | Moderate | Low |
|---|---|---|---|
| High | 0.675 | 0.375 | 0.075 |
| Medium | 0.450 | 0.250 | 0.050 |
| Low | 0.225 | 0.125 | 0.025 |

Table 3. Risk Prioritization Table

| Risk Priority | Rating Low Value | Rating High Value | Risk Planning Level | Risk Plan Location | Risk Review Frequency |
|---|---|---|---|---|---|
| 1 - High | 0.500 | 1.000 | Detailed Risk Plan | Separate Document | At least monthly |
| 2 - Medium | 0.150 | 0.500 | Modest Risk Plan | Risk Register | At least semi-annually |
| 3 - Low | 0.000 | 0.150 | Minimal Risk Plan | Risk Register | At least annually |

# Risk Management (2)

- Most significant risks:
  - Changes in funding levels or delays in funding (i.e., CR)
  - Changes in the rate of technology development, or in the scheduled availability of new hardware components
  - Loss of key project members, due to small number of highly-knowledgeable technical experts working on the project.
  - Loss of system availability due to cyber security incident.

- Most significant impact: inability to meet delivered computing milestones

# Cyber Security Plan

- The system of computing facilities used by LQCD-ext III is classified as a minor application contained in the General Computing Enclave at Fermilab and in the Scientific Computing Enclaves at BNL.

- Security risk assessments, security controls and contingency plans are documented in the security plans for each site, which are prepared in accordance with NIST 800-18, *Guide for Developing Security Plans for Federal Information Systems*.
  - Risk assessments identify possible vulnerabilities and the controls in place to mitigate them.
  - Contingency plans establish procedures to recover systems following a disruption. These plans describe systems, define responsibilities, and establish damage assessment and recovery operations.
  - Security plans are updated for each new system deployment; they are reviewed and signature approval obtained along the line management chain within each host laboratory.

# Cyber Security Plan (2)

- Annual security vulnerability assessments are performed using scanning tools and documentation reviews. Controls are put into place to mitigate any identified vulnerabilities.
- Vulnerability scans are run daily on all production IC computing facilities.
- BNL and FNAL have been certified and accredited (C&A) with Authority to Operate as documented in the LQCD-ext III C&A Document. Copies of ATO documents are maintained by the LQCD-ext III Program Office and included in the C&A document.
- Given that the Site Managers for the extension will be the same individuals overseeing the existing facilities, and given many years of diligent operating experience, we believe the controls and procedures are in place to mitigate known security risks and to quickly respond to any new risks that may arise.

# Quality Assurance

- Quality assurance is addressed in the preliminary PEP and described in detail in the *LQCD-ext III Quality Assurance Plan*.
- Where they exist, established quality control procedures at the host laboratories are followed.
- Additional QA measures:
  - All new hardware is inspected for quality defects upon initial receipt.
  - As new systems are brought online, tests are conducted to verify proper operation at the node and system level.
  - Prior to production release, new systems are released in “user-friendly” mode to stress test the system and identify potential performance defects.
  - System uptime is monitored for all production systems; our metafacility uptime goal is ≥95% .
  - Customer satisfaction is measured through annual user surveys that poll the user community on areas such as overall user satisfaction, system availability, ease of access, quality of documentation, and quality of helpdesk response.

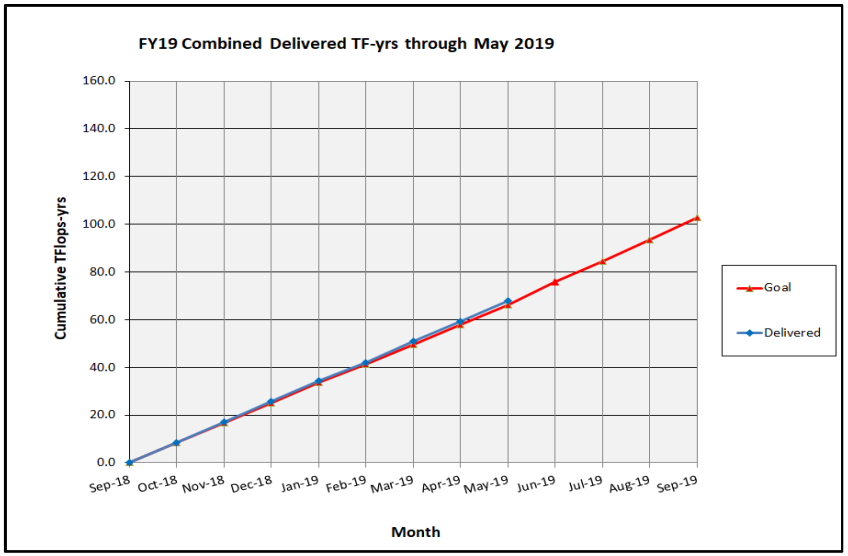
# PERFORMANCE MANAGEMENT

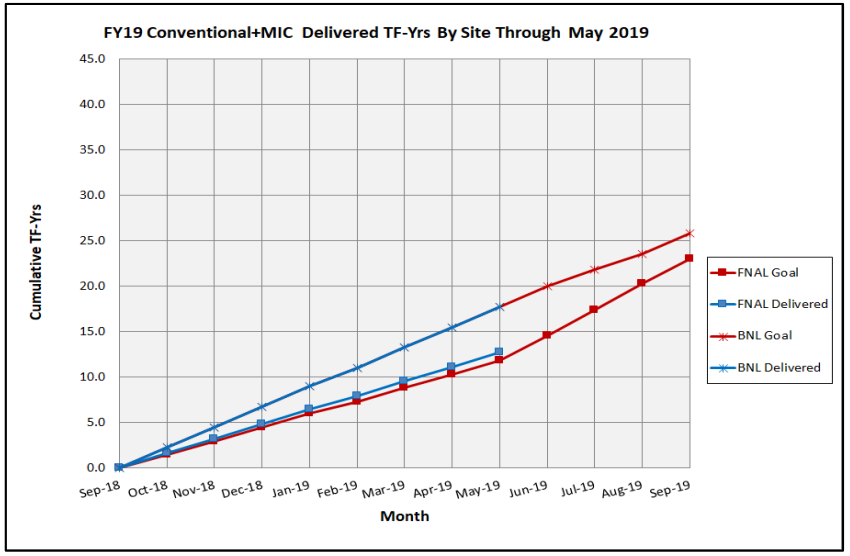
# Performance Indicators and Milestones

- Performance milestones, metrics, and Key Performance Indicators (KPIs) are explicitly defined in the PEP (Table 2, Appendices C and D)
  - 15 Level-1 Milestones (Table 2)
    - Annual computer architecture plan developed and reviewed
    - Annual hardware procurement/deployments
    - Annual aggregate computing delivered
  - 5 Cost and Work Performance Metrics (Appendix C)
    - Annual operations and maintenance
    - Annual procurement and deployment of new systems
  - 24 Computing Facility Key Performance Indicators (Appendix D)
    - Aggregate computing delivered *(TFlops-yrs)*
    - Customer satisfaction *(overall rating ≥ 92%)*
    - Cyber security *(frequency of vulnerability scans = daily)*
    - System performance *(% of average machine uptime ≥95%)*
    - Process improvements *(% of tickets closed within 2 business days ≥95%)*
- Progress against these goals is tracked and reported periodically to the Federal Project Director. Performance is also presented for review at the Annual DOE Progress Review.

# Performance Management Controls

- On a monthly basis, the Associate Program Manager collects performance data from the host sites and prepares graphs showing fiscal-year-to-date aggregate computing delivered by each site, and from the metafacility as a whole.
  - Performance is compared to the baseline plan
  - Deviations are analyzed and corrective actions taken when possible.
- The following graphs are examples from the existing LQCD-ext II project – a similar process will be followed for LQCD-ext III.



# Cost Performance Management

- Cost Controls
  - Site Managers are responsible for executing work in their functional area according to plan and within budget.
  - All procurements >$50K require the approval of the CPM.
  - All procurements abide by host institution procurement rules, procedures, and signature authority requirements.
  - Site managers use host institution accounting and financial systems to manage and track cost performance.

- Cost Management
  - On a monthly basis, Site Managers provide cost and effort data to the Program Office.
  - The Associate Program Manager enters this data into cost and effort tracking spreadsheets and analyzes performance against plan.
  - Tables and charts are generated to compare actuals against baseline plans, and to identify trends that may indicate the need for corrective action.
  - Cost performance is reviewed with the Federal Director on a monthly basis.

.

# Schedule Controls

- On an annual basis, the Associate Program Manager develops the detailed WBS and schedule for the subsequent year, taking input from the Site Managers.
- The WBS and schedule is developed and maintained using MS Project.
- Work activities are updated on a monthly basis
- Progress towards new system deployment milestones is carefully monitored and corrective actions initiated as necessary to ensure that key performance milestones are met.
- Progress on deployment activities are reported to the Federal on a monthly basis.

# Summary

- We plan to operate and manage the extension in a manner like the existing LQCD-ext II project, taking advantage of the experience gained over the past 13+ years of operations and deployments.

- The same management and technical teams will be in place for the extension, ensuring a smooth transition with uninterrupted availability of resources to the USQCD user community. We have been successful in meeting our key performance goals and milestones.

- A preliminary budget plan has been prepared based on funding guidance, and projected performance goals for delivered computing have been developed.

- Both host laboratories have provided letters of support for the extension, noting their commitment for ongoing collaboration and support.