# *LQCD ARRA Computing Project*

## *Overview, Management & Performance*

*Chip Watson*
*Jefferson Lab Scientific Computing Group*

May 16, 2012

# ARRA LQCD Project Definition

The LQCD ARRA Computing Project directly supports the mission of the DOE's Nuclear Physics Program "to foster fundamental research in nuclear physics that will provide new insights and advance our knowledge on the nature of matter and energy...".

The Project also supports the Scientific Strategic Goal within the DOE Strategic Plan to "Provide world-class scientific research capacity needed to: advance the frontiers of knowledge in physical sciences...[and] provide world-class research facilities for the Nation's science enterprise."

The project scope, management structure, and milestones are defined in the Project Execution Plan, a 17 page document submitted after award, and amended to incorporate the evolution in the plans to exploit GPUs.

The technical scope, computing architecture and performance will be addressed in the next talk.

Management, schedule and cost, change control, and operational context will be covered in this presentation.

# ARRA Project Context

The LQCD ARRA project is complementary to the LQCD-ext project

- In 2009 USQCD collaboration requested $24M for "LQCD 2" (5 years)
  - LQCD-ext was funded at $18M
  - LQCD ARRA was funded at $5M, enabling funding of nearly the target amount
- Project co-ordination:
  - Jefferson Lab was to have received the next LQCD cluster in FY 2010
  - A collective decision was made to put LQCD ARRA resource at Jefferson Lab, and to re-locate the FY2010 LQCD-ext machine to Fermilab, shifting it later in the year to create the possibility of a combined FY2010-11 larger machine (which was done)

The LQCD SciDAC project provides the necessary software for both of these computing projects

USQCD proposals for INCITE and NSF allocations address capability computing

The ARRA & LQCD-ext projects target ***high end capacity*** (many jobs < 1 Tflops)

# Project Goals & Phasing

Performance goal: to double USQCD's resources, at that time 17 Tflops.

As an ARRA project, another goal was to move as quickly as possible to get funding into the economy.

- The project was structured as 2 phases
  - Phase 1, $1.78M in hardware to be awarded by the end of FY2009
  - Phase 2, $1.70M in hardware, to follow by ~3 months
- It evolved to include GPUs
  - By the time the project started, it was clear GPUs would be ready for exploitation by LQCD, enabling a significant performance increase
  - Each phase was adjusted to include a GPU component, so that GPU deployment could match the community's uptake of the technology
  - Phase 2 was eventually adjusted to accommodate the timeline for the NVIDIA Fermi GPU
  - In late 2011, most of the 2009 GPUs were upgraded to get a highly cost effective performance boost

As a result of the GPUs, the delivered performance is now effectively 84 Tflops!

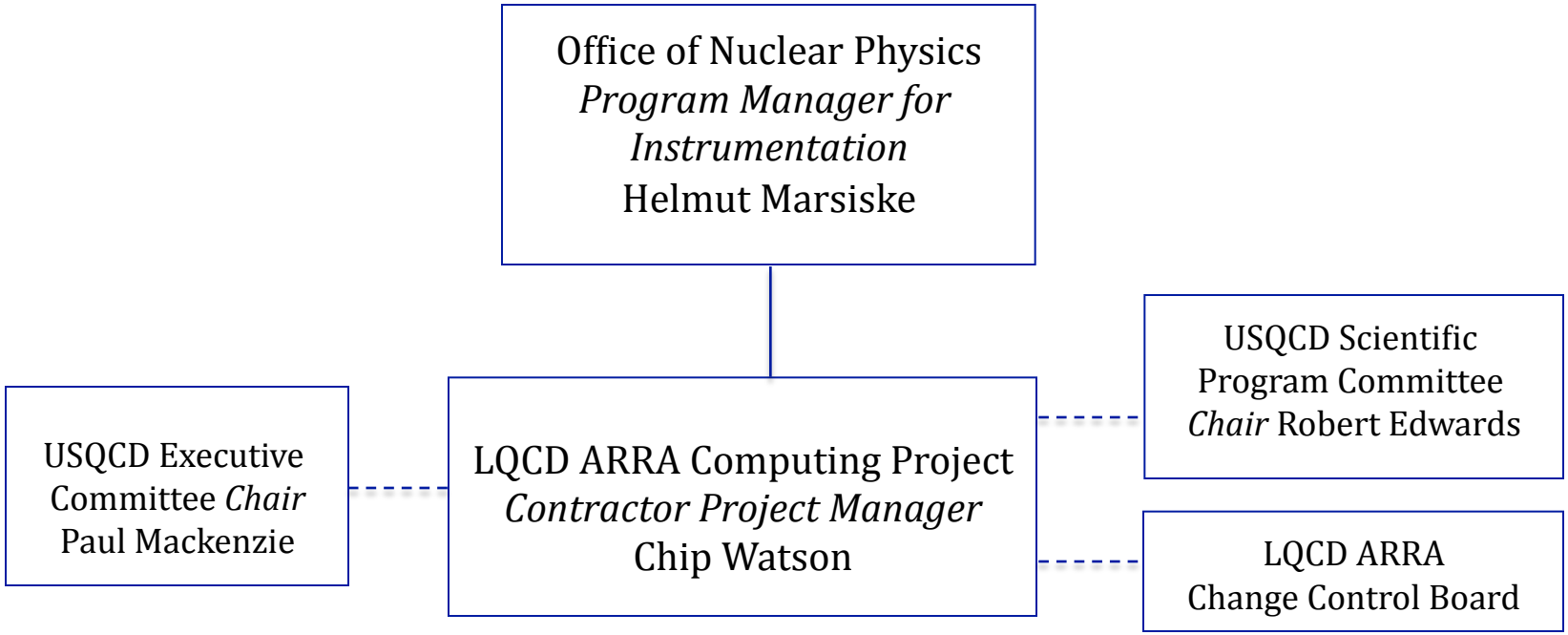
# Management

The ARRA LQCD project is in many ways modeled after the LQCD Computing project, and re-uses the following components or management approaches:

- Relationship to the USQCD Executive Committee
- Relationship to the Scientific Advisory Committee (computing allocations)
- Approach to hardware selection and alternatives analysis, to achieve the greatest performance for dollars invested
- Approach to benchmarking
- Cost model for operations (FTE planning)
- Change control process (but simplified since there is only one site)

Because of the lower total project cost, a single site, and fewer deployment cycles, management is intentionally lighter weight.

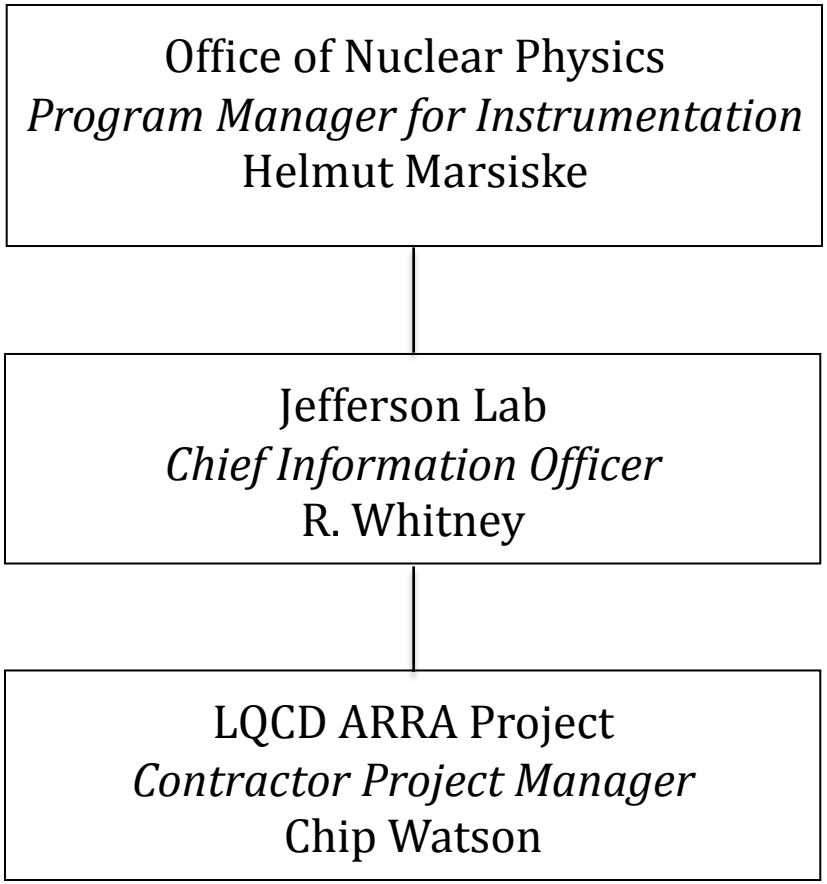
# Management Organization




Management Organization Chart for the LQCD Computing Project.

Vertical lines indicate reporting relationships. Horizontal lines indicate advisory relationships

# Management Authorities



# FY2012 Project Adjustments

A decision has been made to optimize LQCD-ext and ARRA LQCD projects by merging operations at the start of FY2013.

Driver:

ARRA LQCD is finished deployment and is mostly just operations, with a strong overlap in functions with LQCD-ext (operations are the same, just for different hardware).

Implications:

- Transfer out-year operations labor costs to LQCD-ext (FY13 for all ARRA hardware, possibly FY14 for the phase 2 hardware)
- Since all ARRA funds have been received by JLab, spend the remaining funds on FY2012 operations (labor), plus small strategic procurements (last 3% of ARRA LQCD budget)
- A decision on optimal use (accelerated vs non-accelerated cluster augmentation, or disk capacity additions) will be made by mid-July, based upon what provides the greatest benefits to USQCD

# Change Control

| Level | Cost | Schedule | Technical Scope |
|---|---|---|---|
| DOE Program Manager (Level 0) | | > 1-month delay of a Level 1 milestone date | Change of any WBS element that could adversely affect project performance specifications |
| LQCD ARRA CCB (Level 1) | A cumulative increase of more than $200K in WBS Level 2 | > 1-month delay of a Level 1 milestone date | Any deviation from technical deliverables that does not affect expected project performance specifications. |
| LQCD ARRA Contractor Project Manager (Level 2) | Any increase of > $50K in the WBS Level 2 | > 1-month delay of a Level 2 milestone date | Technical design changes that do not impact technical deliverables. |

The GPU upgrade in the Fall was a Level 2 change ($77K).

Merging with LQCD-ext results in a total cost shift of less than $200K for the LQCD ARRA project, but is of a scope sufficient across both projects to be treated as a Level 0 change, and the change control process for both projects will be exercised (ARRA CC committee is a subset of LQCD-ext).

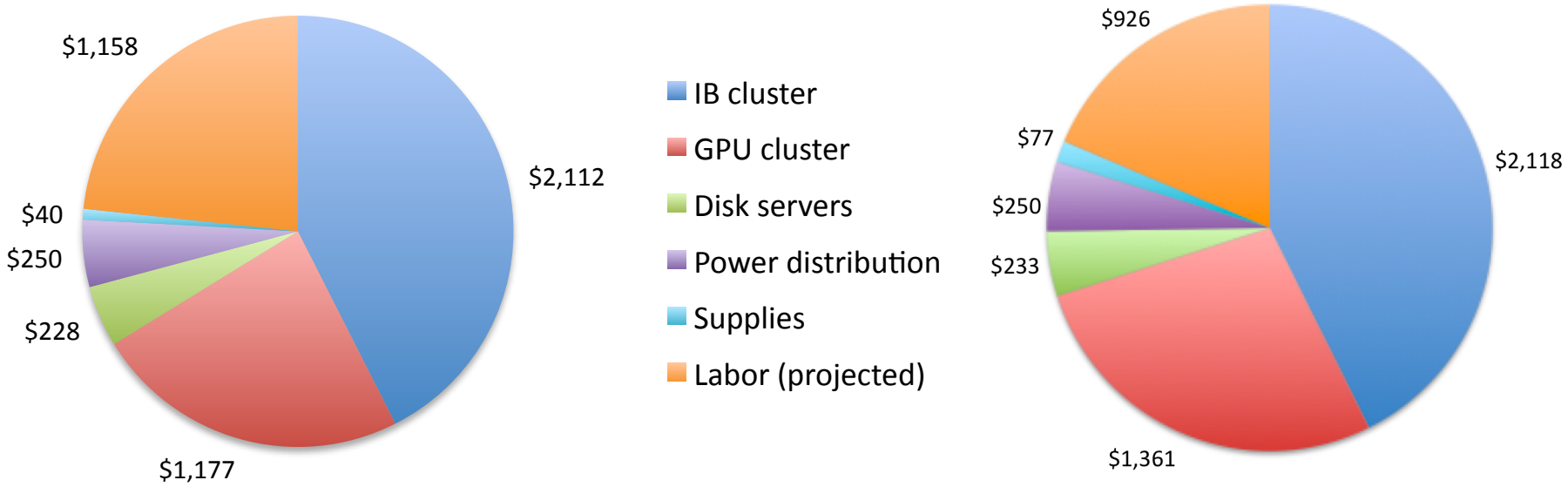
# Budget Plan Adjustments

Original plan (left)

- 70% of the funds for hardware (Infiniband & GPU clusters, disk servers)
- 5% for power conditioning and distribution
- 25% for labor

FY 2012 Update (right)

- GPU upgrades from GTX-285 to GTX-580
- Additional final hardware procurements (tbd)
- Reduced labor costs (1 year shorter operations)




Numbers show are actual costs, with labor projected to the planned end of project.

# Operations

## Summer 2011 Cyber Incident

When the intrusion was detected, Jefferson Lab closed itself off from the internet except for email. Later, white-listed hosts could connect. This happened at worst possible time – just as we were transitioning to a new allocation year at the beginning of July. It was nearly 2 months before we were at anything resembling "normal". Fortunately, on-site users could keep the machine busy and consume their allocations throughout the incident, otherwise cycles would have been lost.

## Staffing Shortage

Compounding our difficulties, one of the four staff of our operations team left for a higher salary with 2 weeks notice right at the start of the incident, slowing down our transition to the new version of CentOS, and our ability to deal with the cyber incident. Computer center was unable to back-fill for us.

## Impact

Disgruntled users, but not much lost running time.

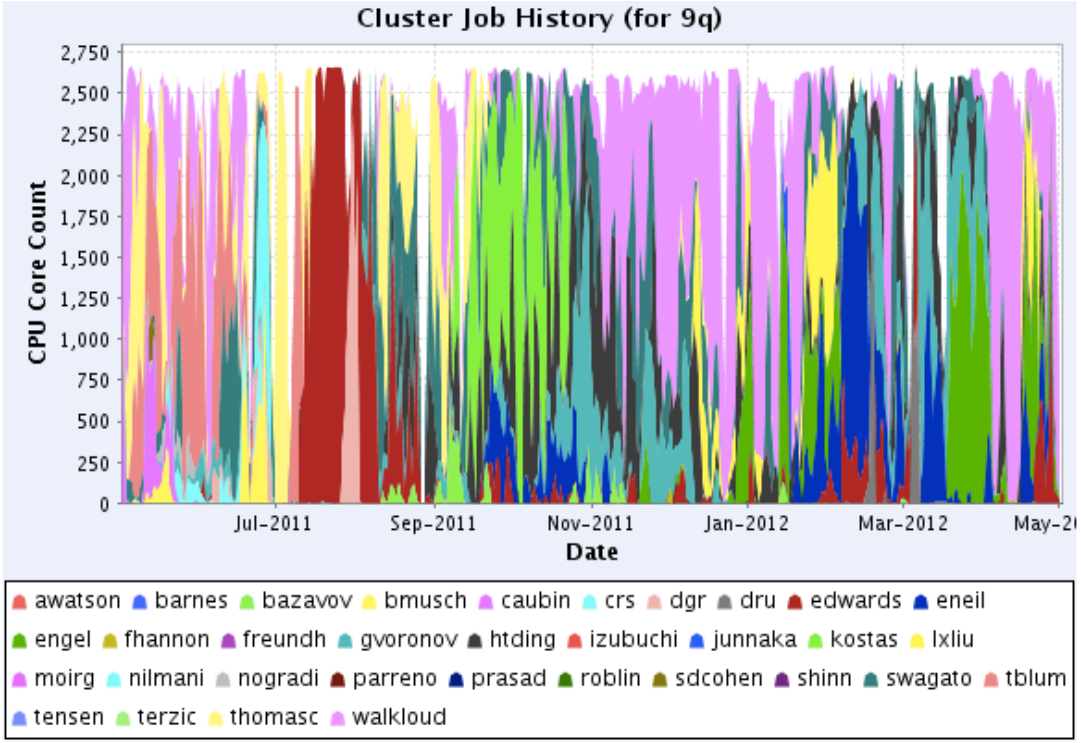
# Infiniband Running Status

Infiniband Clusters:

Cluster runs at > 99% up except for power or cooling disruptions, or system upgrades (e.g. an annual upgrade to a new version of CentOS at the beginning of July for the next allocation year).

Utilization is generally high, with occasional dips due to holidays or conferences, or sometimes all projects catching their breath at the same time. Many of the fluctuations are draining for large jobs to start.

Note that in July one user consumed the cluster throughout the cyber security incident.




Core counts are normalized to 2011-12 standard Fermilab Jpsi core (Opteron) based upon relative performance.

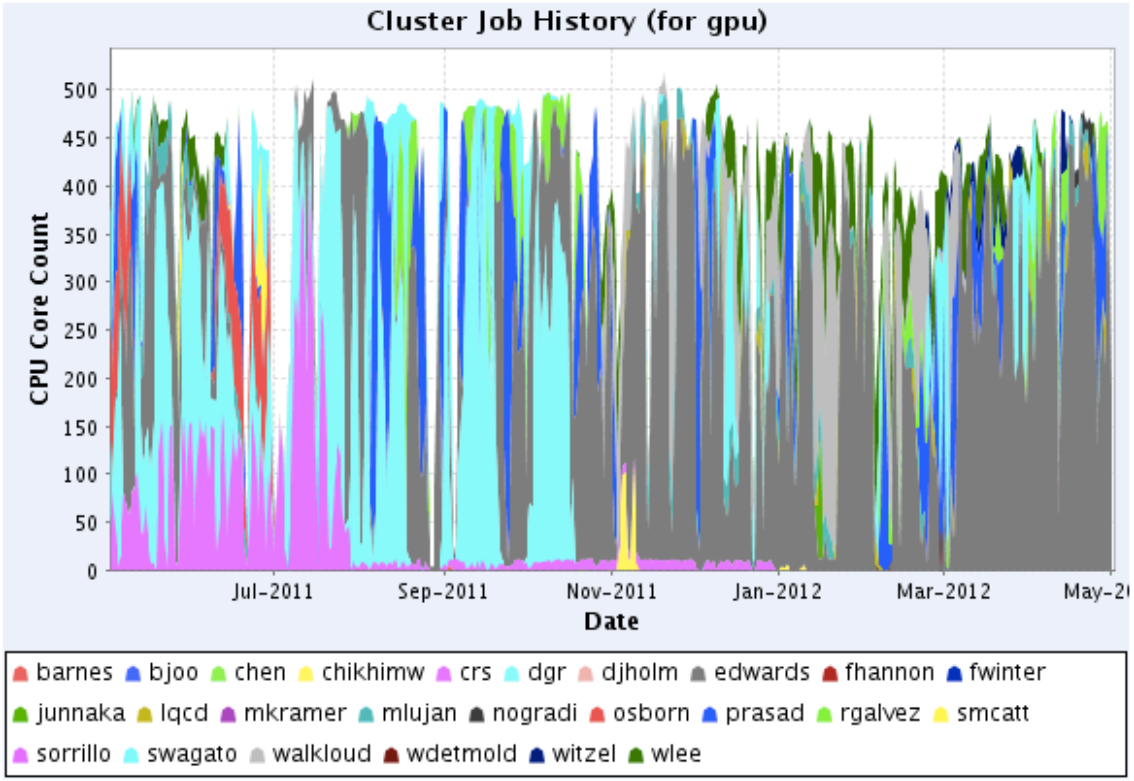
Additional operational charts and graphs can be viewed at http://lqcd.jlab.org/

# Running Status: GPU Nodes

## GPU Cluster

(Graph is in GPU count, un-normalized for performance.)

- The dips in November and January are the upgrades from GTX-285 to GTX-580 cards.
- About 25% of these cards were not acceptable for production (studied in January-March), and many nodes were out of production for a while. Even so, capacity was higher in March than in October due to higher performing cards, and the upgrade was advantageous.
- After weeding out bad cards, some nodes were re-populated with GTX-285 (April-May)



# Running Status: Disk System

## Lustre System

The system consists of

- dual head Meta Data Server
- 23 Object Storage Servers

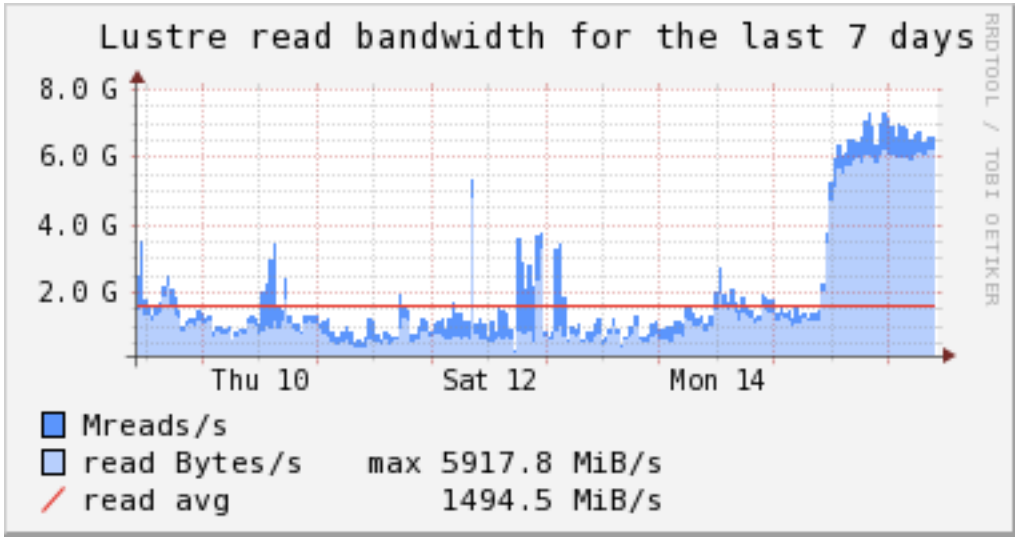
All nodes are on Infiniband, as are all the compute nodes.

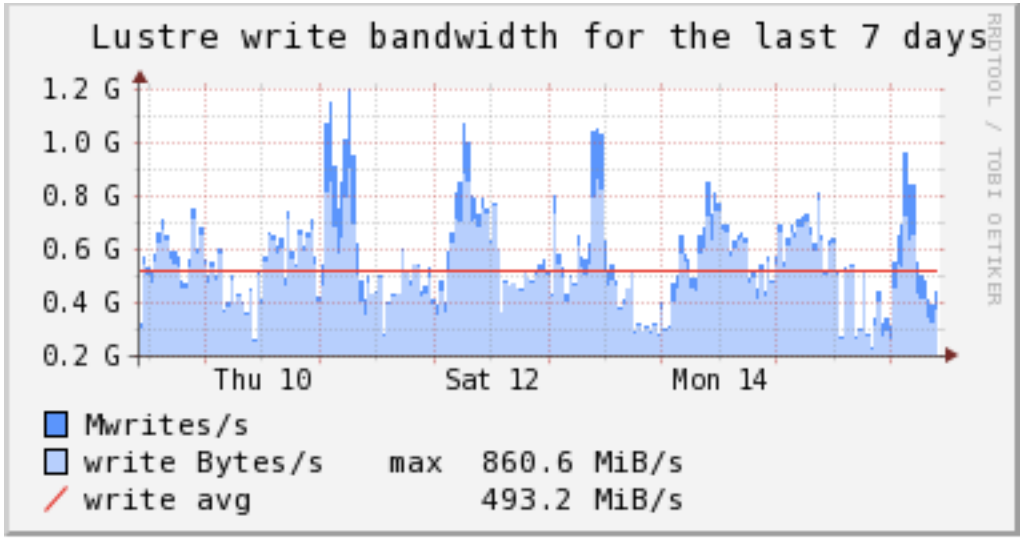
We scale out to increase capacity and bandwidth

- newest systems use 2TB disks
- newest systems have QDR IB

The system performs quite well, with read spikes of up to 8 GB/s, aggregate.

Additional expansions planned for this summer and fall.



# GPUs: A Large Capacity Resource

~500 GPUs in production at Jefferson Lab

- ★ 200,000 GPU cores
- ★ 600 Tflops peak single precision
- ★ **100 Tflops aggregate sustained performance** in the Clover inverter (mixed half / single precision)
- ★ 74 Tflops "effective cluster performance", as measured by production application clock time acceleration compared to non-GPU clusters, weighted by actual usage (correctly accounts for Amdahl's Law)
- ★ Significant increase in dedicated USQCD resources

Additional details in this afternoon's technical talk.

# Summary

The LQCD ARRA project completed its deployment phase on budget and on schedule, is now in stable operations, and will transition all hardware to the LQCD-ext project effective Oct 1.

The project achieved a 5-fold increase in total delivered Tflops capacity (74+10 = 84 Tflops vs 16 Tflops) by moving aggressively to exploit GPUs, while still expanding capacity for non-accelerated codes.