Going Faster

Implementing the S&T Response Plan for ENP & LQCD

Computing Roundtable

June 28, 2018

Chip Watson
Scientific Computing

Outline

FASTER PIPES

BIGGER ENGINE

OUTSIDE REINFORCEMENTS

View from Above

The S&T response in summary

- Jefferson Lab's computing needs as well as data storage requirements (disk and tape) are growing
- 2 fold approach to meeting those needs
  - Increased resources at JLab
  - A growing use of offsite resources
    - OSG
    - NERSC
    - (some day) Clouds

In just a few slides, I will give you some of the near term details of how these things are progressing.

DAQ to Tape

As you might have noticed, things went well this past Spring:

- The accelerator ran more weeks
- CLAS-12 moved towards production running
- Total data rates (4 halls) climbed markedly

You probably also noticed that this (partially unplanned) data rate growth had its downside:

- The tape library's bandwidth was exhausted
- The farm became starved for data (many nodes idle)

Part of this was due to IBM's late delivery to us of LTO tape media for our 4 LTO-8 tape drives (1/3 of our bandwidth), which only started contributing to operations after the run ended.

Improving the Plumbing

FY19 DAQ Target:

- ✧up to 2 GB/s average 4 hall data rate over 24 hours

This number is above 1.2 GB/s requirements so that if transfers are interrupted for a day, there is adequate bandwidth to catch up.

Implications: 6.4 GB/s I/O requirement (during catch-up)

- 2 GB/s from halls
- 4 GB/s to tape (raw + raw duplicate)
- 0.2 GB/s copy sample to Lustre (10%)

Our 2 PB Lustre file system can't easily absorb 6.2 GB/s and still do its primary task of supporting computing (In the Spring, we had to turn off raw-dup, which helped a little).

We need to handle DAQ to tape while also providing

- 1-2 GB/s tape to/from disk
- 2-8 GB/s between Lustre disk and compute

Our Lustre system today can deliver 10 GB/s (only).

Flash Data Buffer

New DAQ gateway from the halls to Scientific Computing

- Separate dedicated flash file server (not in Lustre)
- 20 SSD disks, each 1.9 TB, > 400 MB/s
- 27 TB usable storage, > 6.4 GB/s bandwidth (read + write) (expandable to higher capacity and higher bandwidth)
- RAID-6 (raid-z2) full data protection
- Redundant 40gE connections to campus network (& DAQ)
- Redundant 56g FDR Infiniband connections to tape servers

At 2 GB/s in, disks would fill within just under 4 hours if the tape library failed. The halls will buffer 24 hours, which will be stretched to >72 hours by transfers to Lustre at 1.0 GB/s were absorbing part of the load. Gateway failure also falls back to older 6 * 10g gateways, which can handle the load but degrades Lustre.

During beam off, the flash can be used for random I/O workloads.

Scaling up Storage Bandwidth

Tape Library

- Current total bandwidth LTO-8, -7, -6, -5 (23 drives)

  4*360 + 12*160 + 7*140 = 4.3 GB/s (2.9 during last run)
- Add 4 more LTO-8 tape drives in August = 5.9 GB/s
- Add 4 more LTO-8 tape drives in January = 7.2 GB/s

Lustre File System @ 80% full for better performance

- Currently 0.6 PB for Physics and 0.9 for LQCD
- Adding 0.4 PB for each of Physics and LQCD
- Usable bandwidth at 80% full should grow from 10 GB/s today to 14 GB/s in August

This growth in tape and disk capacity and bandwidth is aimed at supporting both **onsite** and **offsite** computing for FY19.

FY 2019 Lustre Upgrade

In addition to capacity and bandwidth, we are looking to upgrade to a newer version of Lustre with enhanced features:

- Small files' data stored in the metadata server (faster access to the data; no indirection)
- More scalable if needed (can split load onto 2 servers)

As part of this we will upgrade to a new metadata server

- Newer server pair with faster cores (handle higher load)
- Old server pair is already 4 years old

Procedure will take 4-8 months

- Stand up new system (early 2019)
- Migrate data, project by project, to the new system
- Migrate file servers, one by one, to the new system

Computing upgrades on the way

Current system

- 3.5k cores (scaled to Broadwell)

Major farm upgrade due in July

- 88 dual 20 core Skylake compute nodes (farm18)
- adds 3.5k cores (100% gain)

Retiring LQCD cluster to be shared for 6 months

- 250 dual 8 core (2012 Sandy Bridge) compute nodes
- adds 2.4k cores

Size for the Fall run: **9.4k cores** (up 2.7x)

Note: 2.7k cores go end of life mid way through FY2019, and we might add only 1.8k new, dropping onsite capacity to 8.5k cores, still up 150%.

Offsite Computing

Integrate local plus distributed computing models

- Peaks and valleys are becoming larger: LQCD no longer a suitable flywheel to smooth out load variations as they have moved to advanced architectures
- Provisioning to peaks is expensive (idle time wastes money)
- NERSC allocation has been awarded to JLab / GlueX
- New version of SWIF workflow tool which runs atop NERSC is now 80% complete, July beta for GlueX
  - Will yield essentially the same capabilities as SWIF users have now for onsite batch computing.
  - GlueX and soon thereafter LQCD usage planned
  - Will try to add CLAS-12 in FY19 if NERSC can support it
- Still plan to prototype cloud at some point in FY19, to be ready when costs justify it (40% more expensive; interesting for high peak/short duration/high priority jobs)
- Singularity containers are now supported in our farm, OSG, and NERSC for ease of portability

Experimental NP Summary

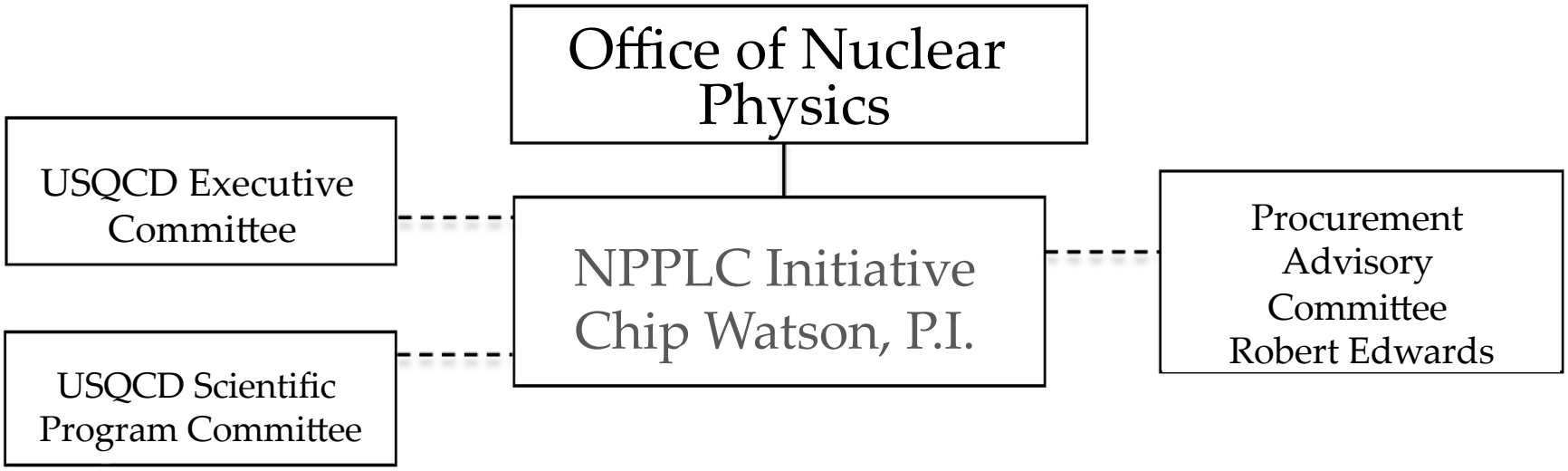
Major growth to support ENP this Fall:

- 100% increase in overall DAQ throughput without limiting offline computing
- 150% growth in tape library bandwidth from Spring run (capacity can grow at $10K/petabyte, no real limits going from current 20 PB up to 200 PB with today's technology)
- 170% growth in ENP peak local computing
- 40% growth in Lustre capacity and bandwidth
- 25% supplement in non-simulation computing by using NERSC. With OSG+NERSC+CLOUD in FY2020, more than 50% of all ENP computing including simulation will be offsite, better supporting large swings in demand (around 3 to 1 peak to valley)

Nuclear and Particle Physics LQCD Computing Initiative

New Activity Summary

- Single lab, NP funded, serves all of USQCD, and is complementary to the modified HEP 2 lab project
- $1M per year, about half hardware, half labor (equals average NP investment per year at JLab for last 10 years)
- Replaces, for JLab, the previous 3 lab, 2 office project (and takes over operations of existing resources at JLab)
- Focus is unchanged: deploy and operate dedicated LQCD optimized resources



- Uses the same standing LQCD advisory bodies (Executive Committee, Scientific Advisory Committee) with a structure identical to the previous LQCD projects, including the 2009-2012 ARRA project
- Reports to the NP Office at DOE, emphasis on NP needs
- Like the ARRA project before it, it should be nearly invisible to the users (i.e. Jefferson Lab remains one of the sites where USQCD does computing)

Quick win for the users: 180 new KNL nodes ! !

The KNL cluster now has multiple partitions

16p partition: 256+4 nodes (2016)

- 32 nodes per switch, 16 uplinks to core (nominally 2:1 oversubscribed, but not really for regular grid problems)
- Xeon Phi 7230 chips, 64 cores at 1.3 GHz
- 192 GB/node, 1 TB disk

18p partitions: 4 single switch mini-clusters

- 44 or 46 nodes per switch, 2 uplinks to core
- Xeon Phi **7250** chips, **68** cores at **1.4 GHz** (faster!)
- 96 GB/node, 150 GB SSD (leaner)

Might be the largest dedicated LQCD resource in the world.

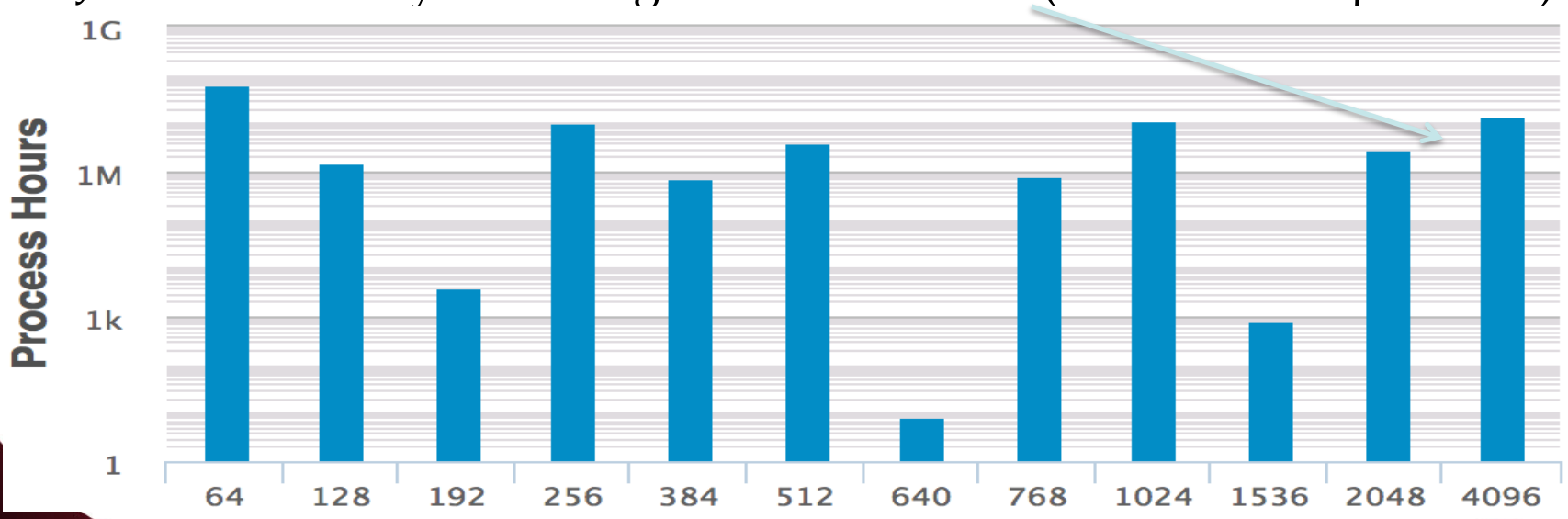
Expanding KNL Resources

Why more 18 month old KNL servers?

- **Only option with 2x performance gain per dollar over last 18 months**
- Easily integrated (experienced ops team, survived initial pain)

Why smaller memory and 4 partitions?

- Memory is 2x as expensive as it was 18 months ago, nodes are much less expensive; as a percent of cost, 192 GB/node was prohibitive
- Network was expensive compared to these very inexpensive nodes
- Job mix has very little usage above 32 nodes (can run in 1st partition)



FY2019 for LQCD

- Operate 444 node KNL cluster
- Support ongoing need for both GPU and non-GPU resources

  Replace JLab's 2012 NVIDIA 45 node Kepler k20m cluster with a newer system, possibly using newest cards:

  quad v100 cards (same as Summit cards, #1 in the world)

  each node 6x as fast as 2012 nodes

  16 nodes? (they are expensive)

  possibly dual 100g network

  100% increase in GPU accelerated capacity

  Exact details to be worked out late summer.

Grand Summary

Scientific Computing at Jefferson Lab is continuing to meet the needs of the laboratory's science.

Working mostly with Physics and Theory, we are adapting to the increasing loads, and to the increasing size of peaks and valleys.