

# JLab “SciPhi-XVI” KNL Cluster



Thomas Jefferson National Accelerator Facility

<https://scicomp.jlab.org/docs/KNLToplevelPage>

Sandy Philpott

[philpott@jlab.org](mailto:philpott@jlab.org)

HEPIX LBNL

October 19, 2016

# Background

## USQCD – DOE LQCD-extII hardware project (2015-2019)

<http://www.usqcd.org/index.html>

3 sites: JLab, FNAL, BNL

in conjunction with USQCD software project

FY16 procurement installed at Jlab in August

~ \$1M, 49TF USQCD benchmarks

Investigated several possibilities ...

- Intel Xeon Phi / Knights Landing
- NVIDIA Pascal GPU, CUDA
- Intel Broadwell CPU server

Consideration factors

- hardware availability timeline
- high speed network – 100 Gbps price/performance; Omni-Path or Infiniband
- available configurations
- **reflective benchmarks of USQCD codes**

**Strategy: optimize portfolio of machines to get the most science on the portfolio of applications**

# *Procurement Process*

Jul '15	Alternatives Analysis & Site Selection
Aug '15	Executive Committee review
Sep '15	FY16 budget finalization
Oct '15	Detailed Acquisition Plan
Nov '15	RFI (done Sep '15)
Mar '16	benchmark suite determination
Apr '16	benchmarks frozen
Apr '16	RFP
June '16	Award

KNL proposals achieved highest aggregate score on all performance metrics and sub-metrics

# Installation Timeline

- May 13: 4 JLab R&D nodes built
- Jun 21: Procurement award announced
- Aug 15-18: Delivery and hardware installation
- Aug 19–30: JLab installation of OS & software  
Benchmarking  
Acceptance testing
- Sep 1: SciPhi-XVI (16p) Early Science running
- Sep 15: JLab becomes an Intel® Parallel Computing Center  
<https://www.jlab.org/news/releases/jefferson-lab-becomes-intel%C2%AE-parallel-computing-center-and-deploys-newest-parallel>
- Oct 1: Production allocations begin (Saturday)
- Oct 3: Upgrade option awarded (1<sup>st</sup> business day of FY17)
- ~Oct 28: Delivery of exercised FY17 upgrade option

# Computing Hardware

## 200 Xeon Phi 7230

64 cores, 1.3GHz, 192GB RAM, 16GB MCDRAM

1 TB disk

Omni-Path

256 threads

4 nodes in 2U

Total: 12,800 cores, 41 TB memory

Jobs run on a dedicated node or nodes

Allocation of 250M Jpsi core hours (FNAL USQCD reference machine)

final adjustment ratio still to be made per final user code benchmarks

Oh! no console capability?! A first for us...

of course, use BMC interface

# *System configuration*

Server Board: Intel S7200AP

BIOS: S72C610.86B.01.01.0190.080520162104 (08/05/2016)

CentOS 7.2, xppsl 1.4.1

(xppsl 1.4.2 available 10/05/2016)

Before this BIOS and xppsl update, fans were at full blast!

Still, hearing protection required if working nearby

# System configuration

- Kernel: 3.10.0-327.22.2.el7.xppsl\_1.4.1.3272.x86\_64
- openMPI:
  - openmpi-1.10.0-10.el7.x86\_64
  - openmpi\_intel\_hfi-1.10.2-8.x86\_64
  - openmpi\_gcc-1.10.2-8.x86\_64
  - openmpi\_gcc\_hfi-1.10.2-8.x86\_64
  - openmpi\_pgi\_hfi-1.10.2-8.x86\_64
- Mvapich2
  - mvapich2\_gcc\_hfi-2.1-4.x86\_64
  - mvapich2\_pgi\_hfi-2.1-4.x86\_64
  - mvapich2\_intel\_hfi-2.1-4.x86\_64
  - mvapich2\_gcc-2.1-4.x86\_64
- Intel Parallel Studio
  - parallel\_studio\_xe\_2016.3.067
  - parallel\_studio\_xe\_2017

# *Memory options*

16 GB MCDRAM (~4x as fast as DDR4)

Flat – user addressable, distinct numa node

Cache – memory side L3

Hybrid – we're not using now

Users can tag their jobs accordingly

We need a tool to be able to change nodes and reboot...

`syscfg`

(`hwloc_dump_hwdata` shows different settings?!)



# Omni-Path Network

Spine : 3 64 port switches; Leaf switch uplinks in bundles of 6,6,4  
for 256 nodes, 2 core switches consumed, 3<sup>rd</sup> support s 4 links to the file system plus 8 additional nodes with a little room to spare

32 nodes connect to a 48 port leaf switch  
- nominally 2:1 oversubscribed, but effectively full bandwidth per node

One core can't drive OPA at full bandwidth

/dev/hfi1 driver – use the correct one, or  
hfi\_userinit: PSM2 and driver version mismatch  
Driver initialization failure on /dev/hfi1 (err=23)  
[3] MPI startup(): tmi fabric is not available and fallback fabric is not enabled

Managed OPA switch subnet manager stopped running several times; recommended for <100 nodes  
so, installed subnet manager on 2 hosts

# Filesystem Services

Nodes need to reach Lustre and legacy NFS on Infiniband...

## LNet routers for Lustre

- a first for us
- Bought 2, added 2 from JLab hosts
- What configuration do we want?
  - 2 procured systems arrived with 2 OPA & 3 IB interfaces
  - **Decided to use 1 OPA and 2 bonded IB**
- Restart each Lustre OSS with new LNet config

## IB / OPA IP routers for NFS

- Configured 2, from JLab hosts

# Web view of 16p

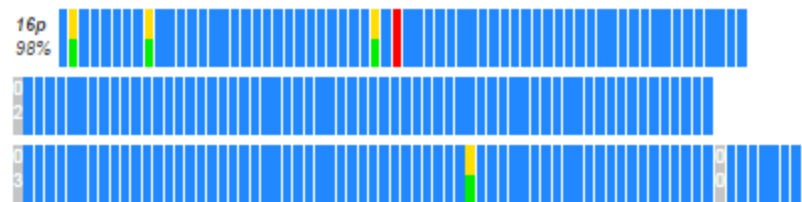
[https://lqcd.jlab.org/lqcd/lqcd\\_index.html#/nodes](https://lqcd.jlab.org/lqcd/lqcd_index.html#/nodes)

[https://lqcd.jlab.org/lqcd/lqcd\\_index.html#/allocation](https://lqcd.jlab.org/lqcd/lqcd_index.html#/allocation)

## LQCD Cluster Node Status

(Click each bar to get individual Node Status Information)

### Phi Clusters



## USQCD Project Allocation Usage (16-17)

(Core hours for each cluster are converted to '16-17' weight unit hours based upon measured relative performance)

Project Name	Allocation*	Project Used Hours**	Annual Pace	Monthly Pace	Remaining Hours	Overused
Spectrump	66,000,000	5,855,300	220%	220%	60,144,699	0
NPLQCDp	20,800,000	0	0%	0%	20,800,000	0
rhqbbarp	2,000,000	82,879	102%	102%	1,917,120	0
cedmpffp	14,000,000	0	0%	0%	14,000,000	0
thermop	31,000,000	6,377,390	510%	510%	24,622,609	0
muon	30,000,000	0	0%	0%	30,000,000	0
Delta	35,000,000	0	0%	0%	35,000,000	0
	<b>198,800,000</b>	<b>12,315,569</b>	<b>154%</b>	<b>154%</b>	<b>186,484,428</b>	<b>0</b>

# Observations

- Double check the power...
  - Our first HPL run tripped the single power panel
  - PDU reached >100 KW, can run at 135KW, 90% of rating, when nodes distributed across 2 panels
- HPL benchmark 329 TF; will rerun with rack 4
- 5 nodes of 200 have hardware errors
  - 3 MCDRAM
  - 1 bad DIMM slot
  - 1 won't POST
- Transparent khugepaged problematic
  - – turned off Oct 18 maintenance day; watching ...

# Still To Do...

- Install 4<sup>th</sup> rack of 64 nodes, including more power circuits in temp location
- Benchmark users' USQCD codes relative to conventional cores
  - Currently assuming 3 Jpsi core hour, the reference for USQCD
- Bin nodes into fast/medium/slow so jobs land on consistently performant nodes (we've seen >10% spread)
- Automate cache/flat memory reboots based on job demands
- Improved automated monitoring, alarms and notification
- Move the cluster to its permanent home in the Data Center
- Upgrade existing clusters to CentOS 7?
  
- Continue to learn and upgrade based on our findings...
- Support more users/projects moving to JLab 16p
- Continue to share experiences with other KNL sites ...

# *HPC BOF – Wednesday 1pm*

Focus on sharing experiences between large sites and smaller sites

- Procurements
- Allocations
- Benchmarks
- Software
- Porting
- Lessons learned