Fermilab Facilities Report

Ken Schumacher

USQCD All-Hands Collaboration Meeting

30 April 2021

Introduction

1. Fermilab Scientific Computing Division
2. NEW Website with Updated Content
3. USQCD "Institutional" Cluster
4. Allocation Year 2020-21 (PY20-21) Usage
5. Progress Against Compute Allocations
6. Storage Update (Disk and Tape)
7. Scheduled Downtimes
8. SLURM Limits and Dispatch Priority
9. User support and SNOW Tickets
10. Questions?

⚛ Fermilab

Fermilab Scientific Computing Division

- The Scientific Computing Division (SCD) provides software solutions and deploys and operates scientific computing facilities in support of the Fermilab program.

- SCD interacts with the experimental community to determine capacity, availability and capability requirements and to procure resources, as necessary.

- SCD engages in R&D activities required to maintain or advance capabilities necessary for the success of Fermilab's future physics program.

- Active Archive Facility (262 PB on tape), CMS Tier-1 Center (24k cores), Data storage and handling, Data Centers, FabrIc for Frontier Experiments (FIFE), General Purpose Grid, HEP Cloud and HPC.

Elizabeth Sexton-Kennedy
Chief Information Officer

James Amundson
Head of SCD

🎔 Fermilab

Updated Web Pages

New web address (WordPress):

 https://computing.fnal.gov/lqcd/

- Automatic redirect from the old web address
- Brings the LQCD pages in line with Fermilab style guidelines
- Also easier for us to update and improve

It is our hope that you can find answers to all your questions on these pages.

We appreciate your feedback so that we can make these pages useful and easy to navigate.

Fermilab USQCD "Institutional" cluster

- 183 nodes, 7,320 cores.
- Intel 6248 "*Cascade Lake*", 2:1 oversubscribed EDR Omni Path.
- 325 GFlops/node - - 59 TFlops total **

- Added 200TB to /lustre1, now 821 TB total
- New Dedicated NFS-server for /project (24 TB) and /home (1 TB) areas.

- Current Type A Allocations total 61.7 M sch (sky-core-hours)
- All Allocation Types total 63.3 M sch

** not applying Amdahl's law.183 nodes,

LQ1

🔹 **Fermilab**

All usage of the LQ1 cluster since July, 1, 2020

Sky-Core-Hours by Project

	total ⌄
— chiqcd	11.420 Mi
— lp3	7.202 Mi
— nme	6.379 Mi
— nedm	6.162 Mi
— vcbok	4.668 Mi
— hadstruc	3.519 Mi
— mslight	3.301 Mi
— rhqbbar	3.241 Mi
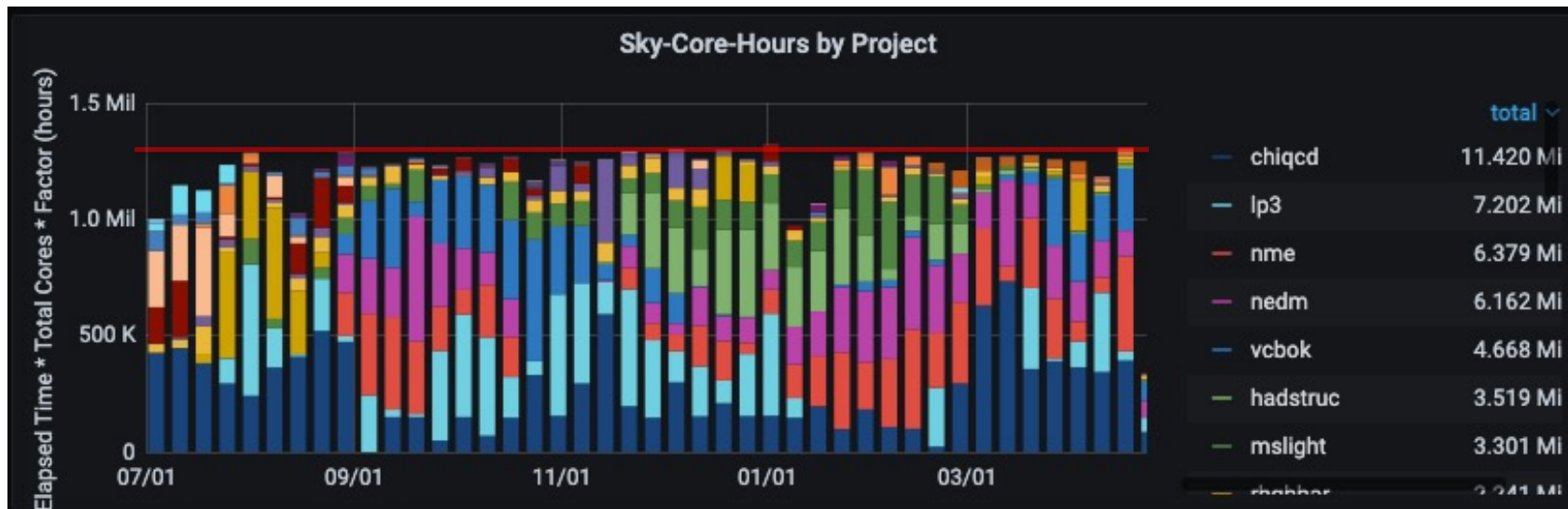
The big dip in early January was an extended downtime for Kernel updates
And the installation of the new dedicated NFS server.

We "Red Line" at 1.25 M sch/week. That's 64.8 M sch/year.

🎇 Fermilab

Just the Type A Allocation projects since July 1, 2020
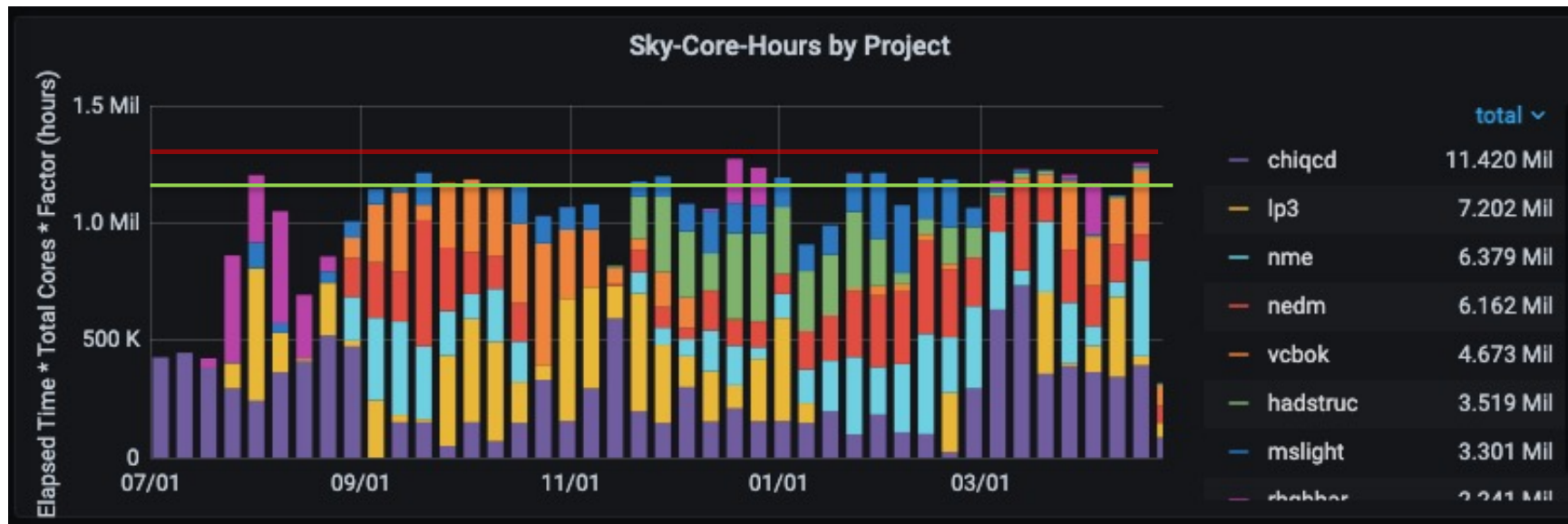
A flat line average to consume the Type A allocations would be 1.19 M sch/week

🐝 Fermilab

Progress against compute allocations as of April 22, 2021

Project Name	Cluster	SPC Original Allocation (Sky-Core-Hours) ▲	Adjustments (Sky-Core-Hours)	SPC Adjusted Allocation (Sky-Core-Hours)	Project Used as of Jul 1, 2020 (Sky-Core-Hours)	Progress against Adjusted Allocation	Remaining Allocation (Sky-Core-Hours)
lp3	FNAL-LQ1	12,500,000	-641,998	11,858,002	7,153,638	60%	4,704,364
nme	FNAL-LQ1	11,400,000	-2,268,859	9,131,141	6,391,211	70%	2,739,930
chiqcd	FNAL-LQ1	11,000,000	1,813,371	12,813,371	11,350,851	89%	1,462,520
nedm	FNAL-LQ1	8,200,000	-171,350	8,028,650	6,155,640	77%	1,873,010
mslight	FNAL-LQ1	6,800,000	-1,151,247	5,648,753	3,297,510	58%	2,351,243
vcbok	FNAL-LQ1	5,000,000	1,355,588	6,355,588	4,604,257	72%	1,751,331
rhqbbar	FNAL-LQ1	3,800,000	-172,408	3,627,592	2,240,849	62%	1,386,743
hadstruc	FNAL-LQ1	3,000,000	1,236,902	4,236,902	3,512,499	83%	724,403
lattsusy	FNAL-LQ1	600,000	-	600,000	550,629	92%	49,371
gluonpdf	FNAL-LQ1	500,000	-	500,000	299,059	60%	200,941

What is not shown here is the hours used by 10 Type C (1.73M) and 5 opportunistic (3.87M) projects

TOTAL	FNAL-LQ1	63,240,000	-1	63,239,999	51,602,652	72.0%	17,682,109

These numbers from April 22 are at 80% through the program year

https://www.usqcd.org/fnal/clusterstatus/lq1/accounting.html

🟦 Fermilab

Storage (disk and tape)

- Please use `lqio.fnal.gov` for large IO transfers to Lustre or tape.
  - Has 100GigE interface to internet.
- Globus endpoint lqcd#lq1 for /lustre1.
- For best job performance, use /scratch
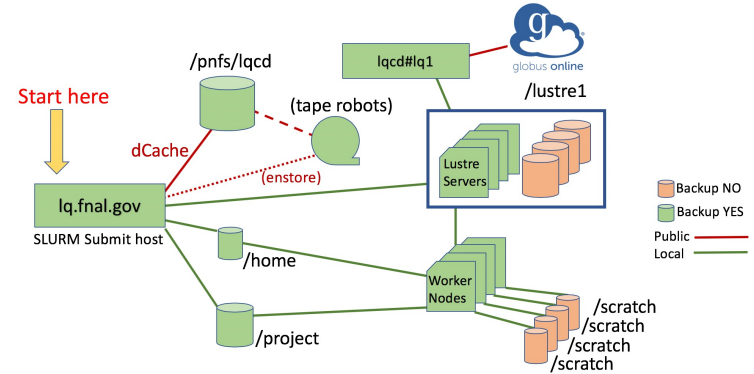
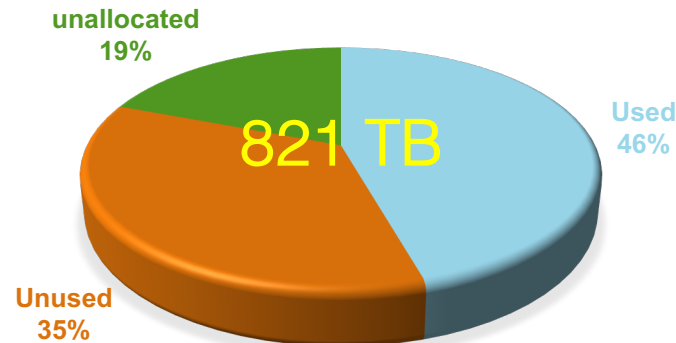| Area	Description	Area	Description
/home	Home area. Visible on all cluster worker nodes via NFS. NOT suitable for configurations or propagators. SUITABLE as a "run" directory for light production or testing. Quota of 5 GB per user. Backups nightly.	/lustre1	Lustre storage. Visible on all cluster worker nodes. NOT suitable for large number of small files. SUITABLE for temporary storage (~month) of very large data files. Disk space usage monitored, and disk quotas enforced. NO backups.
/project	Area typically used for approved projects. Visible on all cluster worker nodes via NFS file-system. NOT suitable for fields e.g. configurations, quark propagators. SUITABLE for output logs, meson correlators and other small data files. Backups nightly.	/pnfs/lqcd	Tape storage. Visible on login head nodes only. NOT suitable for large number of small files, compression highly recommended. SUITABLE for permanent storage of parameter files and results. Must use special copy command: '`dccp`'

https://computing.fnal.gov/lqcd/filesystems/

🐦 **Fermilab**

Usage of /lustre1 disk allocations

Project Name	Filesystem	used (TB)	quota (TB) ▲	limit (TB)	grace	files
rhqbbar	/lustre1	105.63	200.0	200.0	-	8,635,063
chiqcd	/lustre1	54.66	170.0	170.0	-	3,404
lqcd	/lustre1	82.22	95.0	95.0	-	11,125,743
heavylight	/lustre1	36.63	39.0	39.0	-	442,375
fourpluseight	/lustre1	20.59	27.0	27.0	-	17,966,764
higgs	/lustre1	23.00	23.0	23.0	-	673,431
stgmugm2	/lustre1	7.84	20.0	20.0	-	108,313
tumqcd	/lustre1	0.64	20.0	20.0	-	10
vcbok	/lustre1	12.04	15.0	15.0	-	25,765,383
nedm	/lustre1	10.26	15.0	15.0	-	437,695
hadstruc	/lustre1	6.84	10.0	10.0	-	202,070
g4	/lustre1	2.04	10.0	10.0	-	470,211
lp3	/lustre1	4.82	5.0	5.0	-	1,629,028
nme	/lustre1	0.69	5.0	5.0	-	540,539
mslight	/lustre1	2.77	2.0	2.0	-	17,593

What is not shown here are 14 projects with allocations of 1 TB or less. Those 14 are using just 3.6TB.

TOTAL	-	374.23	663.4	663.4	-	72,156,154

https://www.usqcd.org/fnal/clusterstatus/lq1/diskusage.html

LUSTRE SPACE

unallocated 19%
821 TB
Used 46%
Unused 35%

- Each project should have a plan for managing their data over its lifetime.
- Any data that remains three months after the end of the program year is subject to deletion by LQCD-Admin staff, unless an extension has been negotiated.

https://computing.fnal.gov/lqcd/fermilab-data-management-guidelines-policies/

🎇 Fermilab

Usage of /project disk allocations

Project Name	Filesystem	used (GB)	quota (GB) ▲
fermilab	/project	11,217.5	15,360.0
rhqbbar	/project	1,036.8	1,126.4
VcbOK	/project	791.8	1,024.0
hpqcd	/project	849.7	850.0
gluonpdf	/project	0.0	500.0
Geant4	/project	36.0	500.0
rbcdwf	/project	89.9	90.0
fourpluseight	/project	68.2	70.0
hadstruc	/project	0.0	50.0
nonperturbativestaggered	/project	18.2	18.5
mixbk	/project	7.6	8.0
Baryon	/project	5.7	6.0
s1080	/project	0.0	5.0
ok-action	/project	0.0	5.0
hadtensor	/project	3.0	5.0
charmonium	/project	0.3	5.0
c51	/project	1.6	5.0
TOTAL	-	**14,126.3**	**19,627.9**

PROJECT SPACE
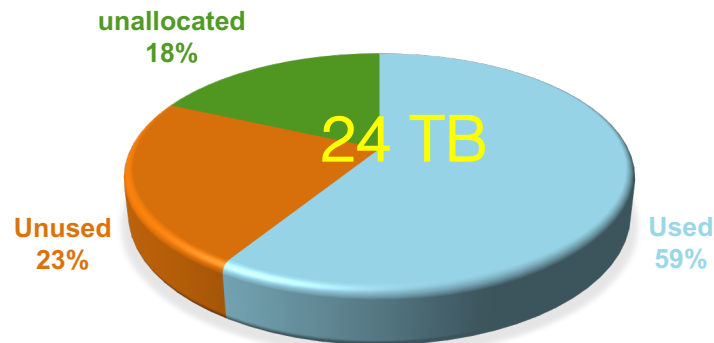
- The large "fermilab" allocation under /project is shared by local LQCD users and IC users outside of LQCD.
- The Project areas must also be cleared when the allocation ends

https://www.usqcd.org/fnal/clusterstatus/lq1/projectusage.html

🎇 **Fermilab**

Scheduled downtimes

- We plan for two downtimes per year for Kernel updates and other Security patches.
    - Early July during transition to new Allocations
    - Early January as the 6 month mark.

- When we need to do other scheduled maintenance, it will typically be on the third Wednesday of the month.

- We will continue to use the lqcd-users@fnal.gov mailing list to announce both scheduled downtimes and unscheduled outages.

- The scheduled maintenance window for Enstore and dCache is the third Wednesday of each month, beginning at 8am, normally 4-8 hours.

🟁 **Fermilab**

SLURM Limits and Dispatch Priority

- Maximum number of nodes per job and nodes per account is 64.
- Maximum number of jobs submitted per account is 125.

QoS Name	QoS Description	Priority	Group Resource Limit	Maximum Wall-time	Maximum Jobs Per User	Maximum Job Submits Per Account
admin	Reserved for Administrators	600				
test	Quick turnaround testing	500	cpu=80 (node=2)	00:30:00	1	3
normal	Normal QoS (default)	300		24:00:00		125
opp	Opportunistic	0		08:00:00		125

Job's requesting shorter wall-time limits are more likely to start sooner as Backfill jobs.

https://computing.fnal.gov/lqcd/slurm/ and https://computing.fnal.gov/lqcd/job-dispatch-explained/

🟥 Fermilab

User support and SNOW Tickets

- All emails to lqcd-admin@fnal.gov will automatically generate INC tickets.
- Ticket Types in SNOW (Service Now)
  - INC – Incidents are for when something is broken. Higher priority and tracked closely
  - RITM – Requests are for all other questions or assistance. Still tracked but not as much reporting up to management or DOE
- We will convert INC to RITM when appropriate.
- Your email reply becomes part of the worklog of the ticket number in the subject

- Emailing lqcd-admin@fnal.gov is the best way to reach us and will elicit the quickest response. Please <u>do not</u> email any of our staff directly.
- Response time as well as time to resolve are tracked and reported
- Our goal is to resolve your issue within 3 business days.

https://computing.fnal.gov/lqcd/support/

🟦 **Fermilab**

Questions ?

I know I just asked you to use the lqcd-admin@fnal.gov email address, but if you have questions about this presentation, you are welcome to contact me:

Ken Schumacher <kschu@fnal.gov>

🟦 Fermilab

The first baby bison of 2021 arrived Monday morning (April 26).

Ken Schumacher l 2021 USQCD All-Hands Collaboration Meeting

🔷 Fermilab