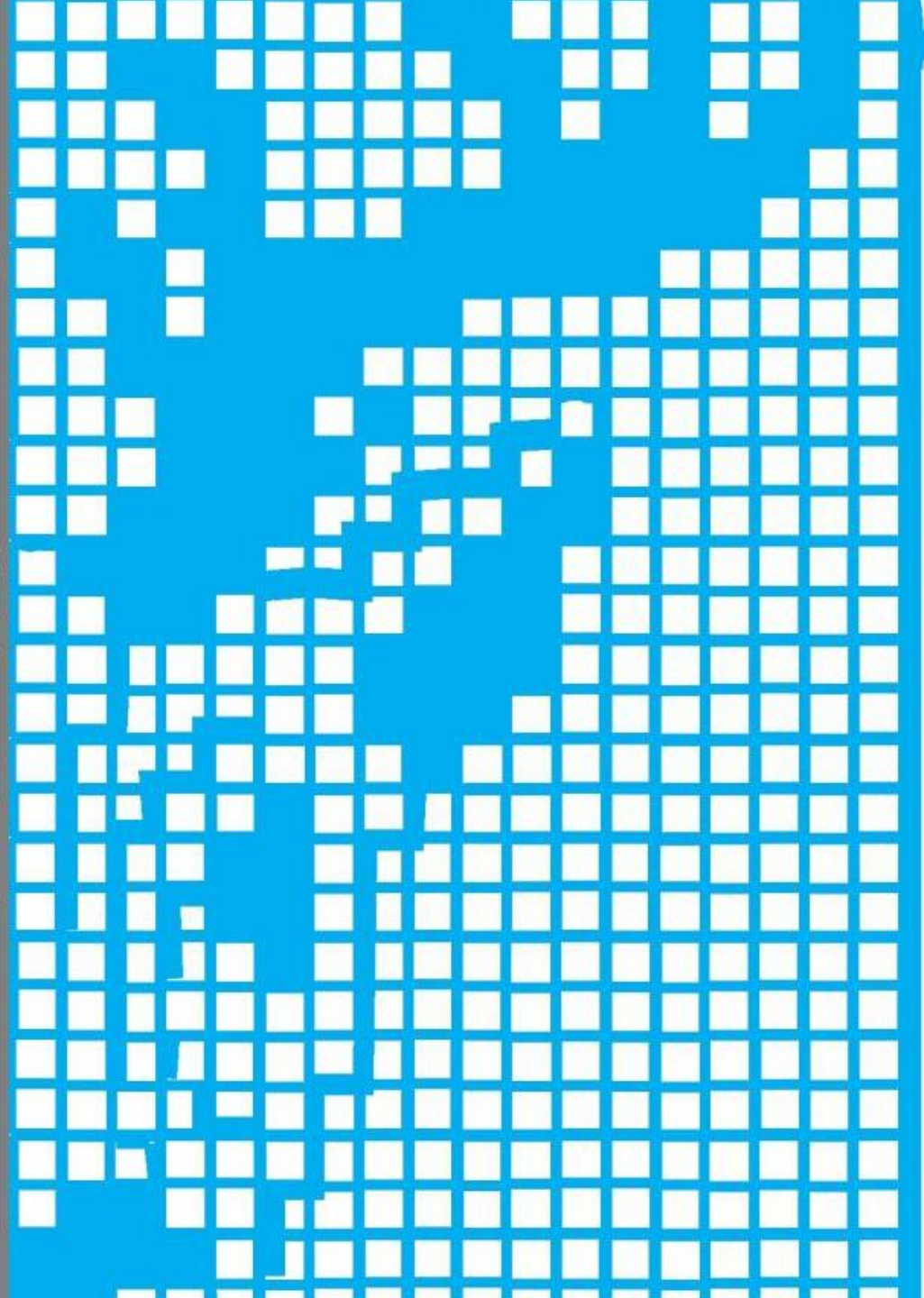# Site Report HPC2N

NDGF All Hands Ljubljana

2024-09-26

Status WLCG compute – Aigert (aka g-ce01)

- 31 PowerEdge R6625, 256/cores, 3GB/core, 3.2TB/node
  - 25.4 HS23/core (~6510/node)

- In production since late May,

- Kebnekaise(-ce) is gone (mostly, lingers on as a test VM)

- Hardware seems to be working fine
  - Personally unsure if network is enough
  - 25Gb/256 cores < 100Mb/core
  - 4×100Gb / 31×256 job slots ~ 50Mb/job slot

- Software side on the other hand …

ARC-ish problems - infosys

- Infosys getting out of sync
  - Noticed that the number in infosys was not matching what slurm AND ARC itself thought about the number of jobs
    - Jobs themselves were even missing
  - Restart helped for a while but then it quickly went out of sync again
  - BDII (really newer openldap) had default limit for 10MB for max size of ldif file (thanks Kildetoft)
  - Manually patched now

ARC-ish problems - Datastaging

- Datastaging "stalling"?

  - Got serious the week after RO started his vacation!

  - Leads to lots of cores getting unused and some job failures (timeout)
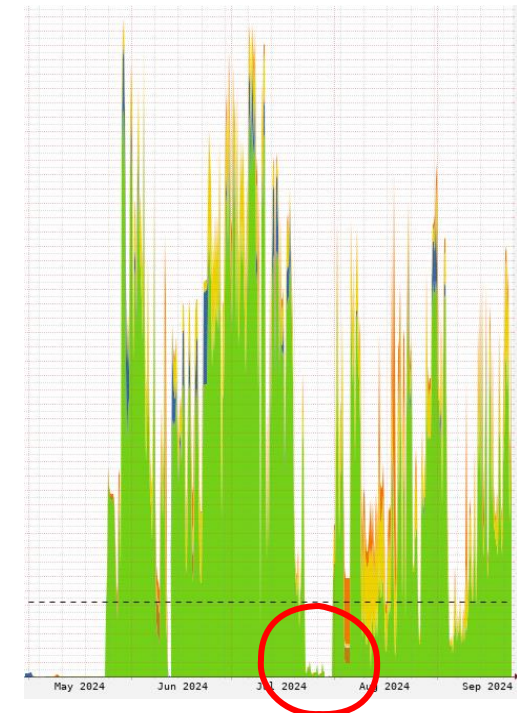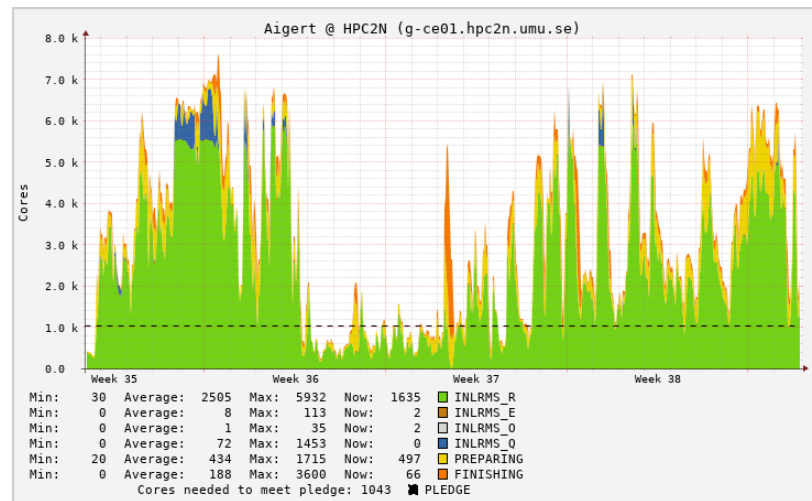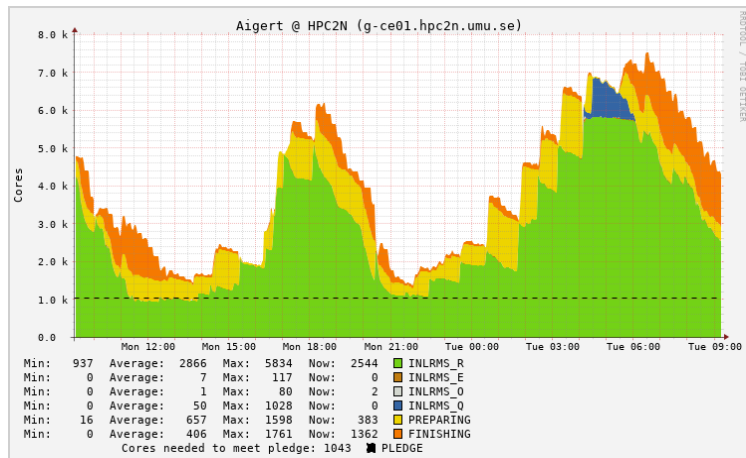
ARC-ish problems – Datastaging (2)

- Problem seems to be that transfers on the CE seems to wait on nothing

  - arcctl says 101 transfers, on actual node zero, nada
  - The logs says the jobs have finished on the datastager machine and the CE is waiting on …. nothing?
  - Nordugrid bug 4191

```
root@g-ce01:~# arcctl  datastaging dtr
Number of current datastaging processes (files):
    State                 Data-delivery host  Number
    CACHE_WAIT            N/A                      154
    TRANSFER             N/A                      900
    TRANSFERRING         q-h39-s.hpc2n.umu.se     101
    TRANSFERRING         q-h36-s.hpc2n.umu.se      88
    TRANSFERRING         q-h38-s.hpc2n.umu.se      94
    NEW                  N/A                        4
----------------------------------------------------------
    TRANSFERRING TOTAL   N/A                      283
    ARC_STAGING_TOTAL    N/A                     1341
```

```
root@g-ce01:~# mssh -n Q_M 'pgrep -c DataStag'
q-h36: 0
q-h37: 0
q-h38: 0
q-h39: 0
```

```
[2024-08-22 10:44:12] [INFO] [277501/3] DTR 25f6...36b8: Transfer finished: 43062 bytes transferred : checksum adler32:14b15d2d
[2024-08-22 10:44:12] [VERBOSE] [277501/3] DTR 25f6...36b8: TRANSFERRING->TRANSFERRED
[2024-08-22 11:04:04] [INFO] [30002/4] DTR 25f6...36b8: Transfer finished: 43062 bytes transferred : checksum adler32:14b15d2d
[2024-08-22 11:04:04] [VERBOSE] [30002/4] DTR 25f6...36b8: TRANSFERRING->TRANSFERRED
```

ARC-ish problems – CA certificates?

- Problems downloading pilot3-dev.tar.gz

- Noticeable with arcctl as transfers in state NEW

- As far as we can tell it seems to be a certificate issue … in Manchester … or maybe Cern?

```
[2024-09-24 09:18:45] [ERROR] [381966/3859863] DTR 90fd...d22d:
Failed to establish connection: TLS: GENERIC_ERROR (Certificate
/C=GB/ST=Greater Manchester/L=Salford/O=Sectigo
Limited/CN=Sectigo RSA Organization Validation Secure Server CA
failed Globus signing policy)
[2024-09-24 09:18:45] [ERROR] [381966/3859863] DTR 90fd...d22d:
Failed checking source replica http://cern.ch:80/atlas-panda-
pilot/pilot3-dev.tar.gz: Failed to obtain information about file:
Certificate /C=GB/ST=Greater Manchester/L=Salford/O=Sectigo
Limited/CN=Sectigo RSA Organization Validation Secure Server CA
failed Globus signing policy
```

```
root@g-ce01:~# arcctl datastaging dtr
Number of current datastaging processes (files):
State                          Data-delivery host   Number
CACHE_WAIT                     N/A
TRANSFER                       N/A                   154
TRANSFERRING                   q-h39-s.hpc2n.umu.se  900
TRANSFERRING                   q-h36-s.hpc2n.umu.se  101
TRANSFERRING                   q-h38-s.hpc2n.umu.se  88
NEW                            N/A                   94
                                                      4
TRANSFERRING TOTAL             N/A
ARC_STAGING_TOTAL              N/A                   283
                                                     1341
```

ARC 7 and ARC cache

- As attempt to debug/fix the datastaging issue, updated one ARC cache machine to ARC 7 (and also jammy)
  - o (The other ones are ARC6 and focal)

- Seen no visible change (neither good or bad)
  - o But at least it has been tried with ARC7 and ARC6 compatibity.

- Only thing was a need to go back to NFS 4 to avoid some NFS hangs

- And since we are on ARC cache, we should mention q-h36 ...

The saga of troublesome ARC cache machine

Finally fixed!

Network

- UMU (finally) moving towards a 100G based core network
    - Eventually
- HPC2N (finally) will get its 100G core switch
    - With 100G uplink
    - Aaany year now
- LHCOPN 100G
    - In production since autumn (late September 2023)
    - Now with spoof filter enabled in the SUNET routers.... Oops.
    - Planning on moving to a 2x100G trunk to have less panic if there's fibre/optics issues, going slow due to UMU net staff shortage

Tape/backup

- Same as last time
- IBM TS4500 library (2550 slot capacity)
- 6x TS1155 tape drives (JD tapes, 15T, 360 MB/s)
- 6x TS1170 tape drives (JF tapes, 50T, 400 MB/s)
- Dell R750
  - 2x100G Ethernet
  - 4x32G FC
  - Approx 30T NVMe for DB and incoming stgpool
  - A few TB of SAS SSD for log mirrors etc
  - Approx 250T spinning disk for on-disk backup storage
  - 256 G RAM, 2xIntel Gold 6334 (total 16 cores @ 3.6 GHz)

Tape pools

- High CPU load when doing 1.4+1.4 GB/s
- Size new pools with this in mind, or assume that Darren will fix the CPU-eating? :-)
  - Looks like we'll need to go from 8-core to 16-core if we want more CPU...

```
top - 11:19:53 up 18 days, 22:01,  1 user,  load average: 59.61, 55.52, 54.19
Tasks: 310 total,   1 running, 309 sleeping,   0 stopped,   0 zombie
%Cpu(s): 40.7 us, 43.0 sy,  0.0 ni,  1.0 id,  0.2 wa,  0.0 hi, 15.1 si,  0.0 st
MiB Mem :  46794.5 total,   2176.4 free,   9735.6 used,  34882.4 buff/cache
MiB Swap:    128.0 total,      0.0 free,    128.0 used.  32901.2 avail Mem


   PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
  1866 globus    20   0 8503308   2.7g     32 S 619.7   5.9  36706:53 java
251260 globus    20   0  382692  45140  35616 S  35.7   0.1   1:27.37 dsmc
251310 globus    20   0  382700  44980  35380 S  32.0   0.1   0:54.94 dsmc
251235 globus    20   0  383688  46116  35820 S  27.7   0.1   1:34.76 dsmc
```

Tape pools (2)

- Procuring new ones as we speak, discussion/offers from vendors
- CPU performance vs TSM per-core licensing dilemma
  - Current suggestions are AMD EPYC 9124 3.0GHz, 16-core
  - Or Intel equivalent
  - Intel has 8-core CPUs in same generation, AMD:s 8-core are previous-gen?
- Looking at around 30 PB write endurance per pool
  - 5x3.84 TB ReadIntensive NVMe with hardware RAID5
- Staying at 25GbE, but aiming to be able to fill it bandwidth-wise this time...

UMU compute - Kebnekaise

- 62 old nodes (1896 cores) + 25 new nodes (3024 zen4 cores)

  - Old nodes zen3 (272) and skylake (1624 cores) (broadwell gone)

  - Zen4 between 48, 64, 96, and 256 cores / node

  - Memory at least 2516 MB/core (6, 8, 10 GB)

  - 14 of the new nodes have new GPUs (32 L40s, 8 H100 SXM5)

  - 3 new nodes use old GPUs (16 A40, 2 A6000)

  - 13 old GPU nodes (20 V100, 4 A100, 2 MI100)

- Many partitions, but auto-selected by some lua-pluginscript

Disclaimer! Some data might be inaccurate

Staff

- Niklas Edmundsson – Tape storage
- Mattias Wadenstein – NEIC NT1
- Erik Andersson – Cluster/Storage/SSC
- My ~~Karlsson~~ Ekman – SUPR/SAMS
- Roger Oscarsson – ARC/WLCG
- Paul Dulaud – DDLS developer
- Abdullah Aziz – DDLS developer
- **Nalina Hamsaiyni Venkatesh** – DDLS developer

- ~~Lars Viklund – App Expert~~
- Åke Sandgren – App Expert
- Birgitte Brydsö - App Expert
- Pedro Ojeda May – App Expert
- Björn Torkelsson – Tech director
- Lena Hellman – Administrative asst.
- Mikael Rännar - WLCG
- Jerry Eriksson – Advanced Consultant
- Paolo Bientinesi - Director