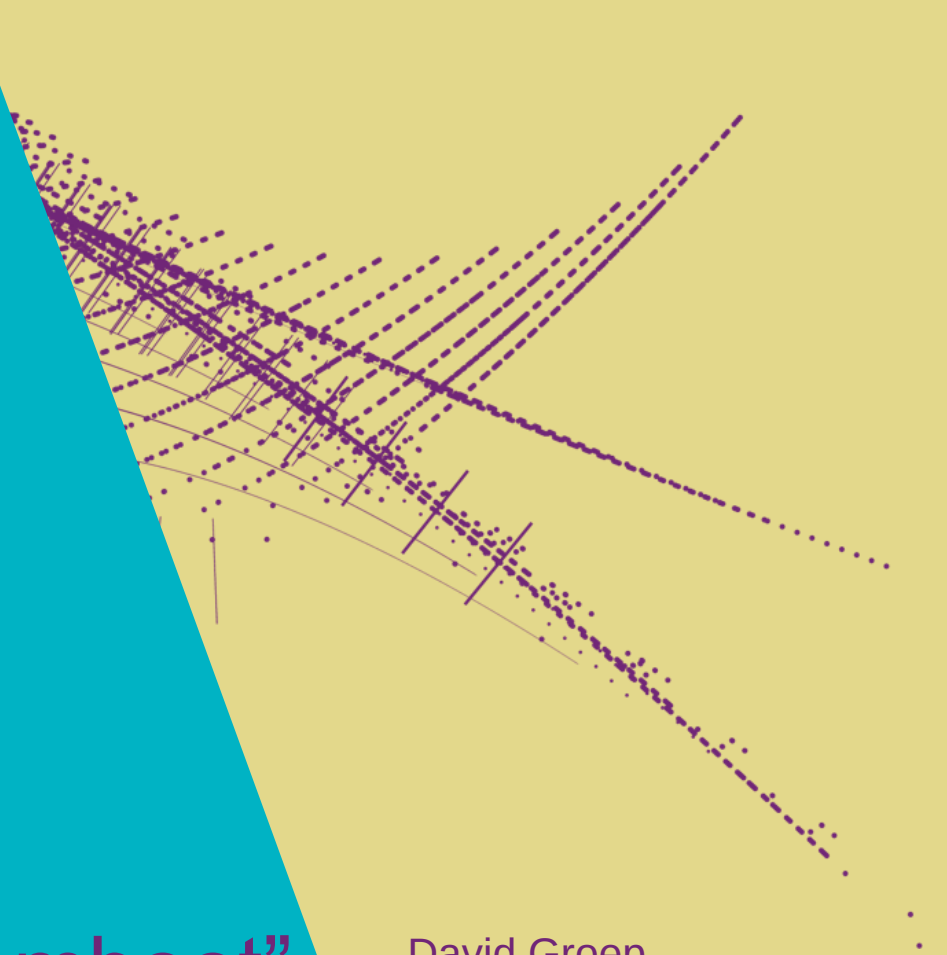
Nikhef

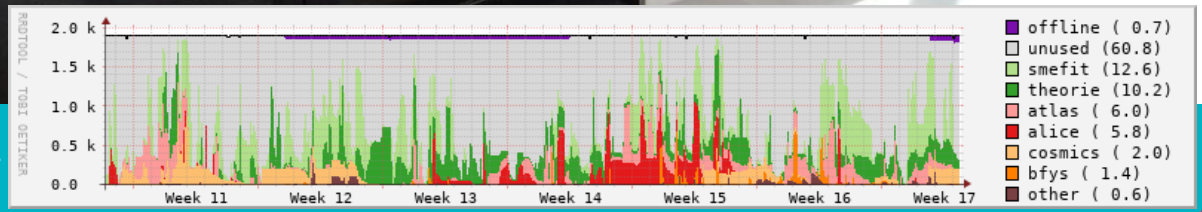


PDP – Physics Data Processing

See Santa's Ship Sail again !

"Zie ginds weer de stoomboot"

David Groep
Nikhef Jamboree 2024



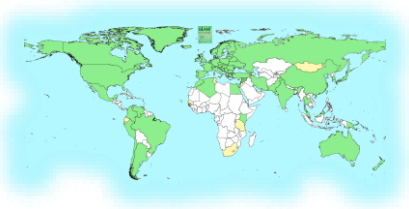
Zie ginds weer

Nikhef

‘Pillars’ of Nikhef Physics Data Processing

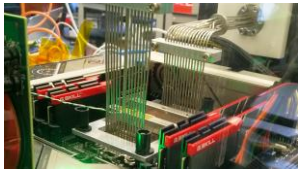
Infrastructure for trusted collaboration

- trust and identity (‘SSO’) for enabling communities
- managing complexity of collaboration mechanisms
- securing the infrastructure of our open science cloud



Algorithmic design patterns and software

- designing software for (GPU) accelerators, new algorithms, high-performance processors
- software design patterns for workflow & data orchestration

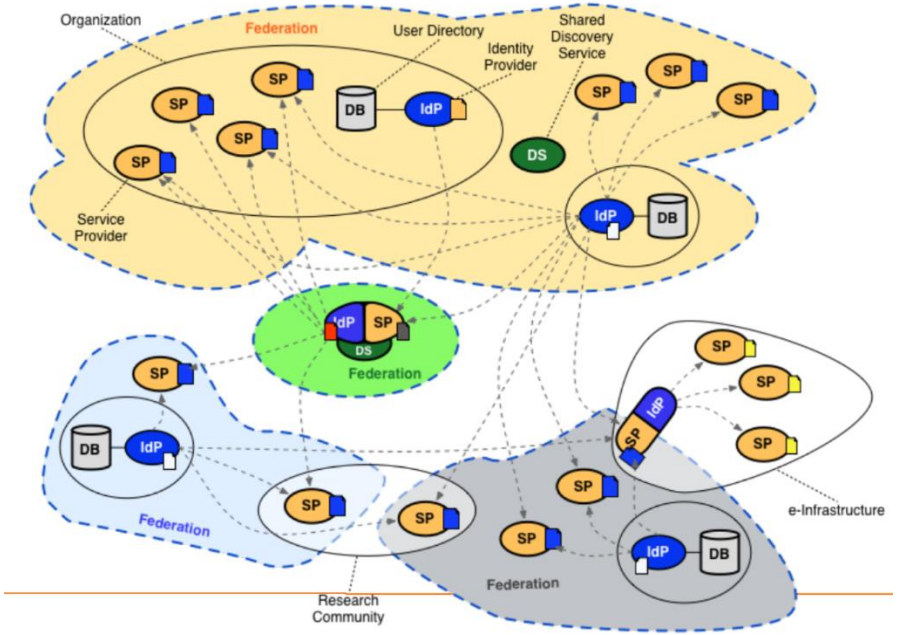


Infrastructure, network & systems co-design R&D

- building ‘research IT facilities’
- co-design & development
- big data science innovation
- research *on* IT infrastructure

Zie ginds weer de stoomboot (see Santa's ship sail again)

Head, heart, and hands: infrastructure for collaboration



multi-federation image by Lukas Hammerle, SWITCH

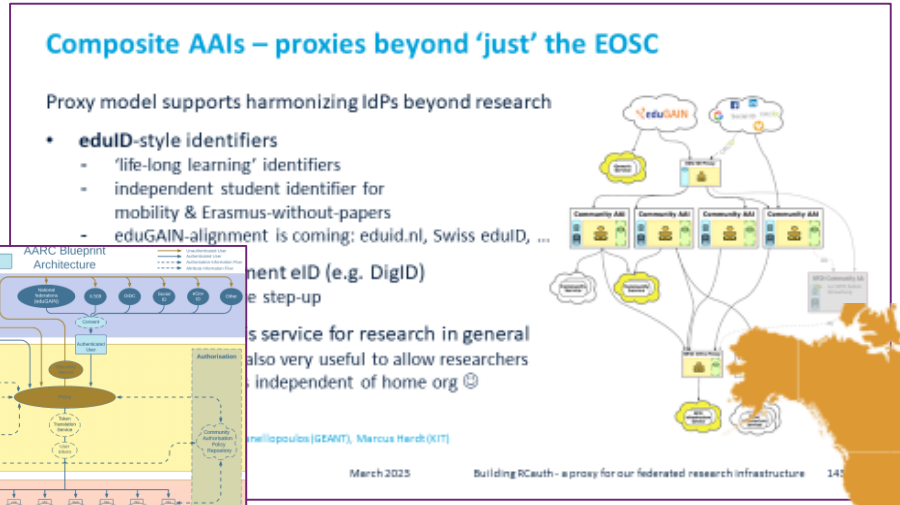
Zie ginds weer de stoomboot (see Santa's ship sail again)

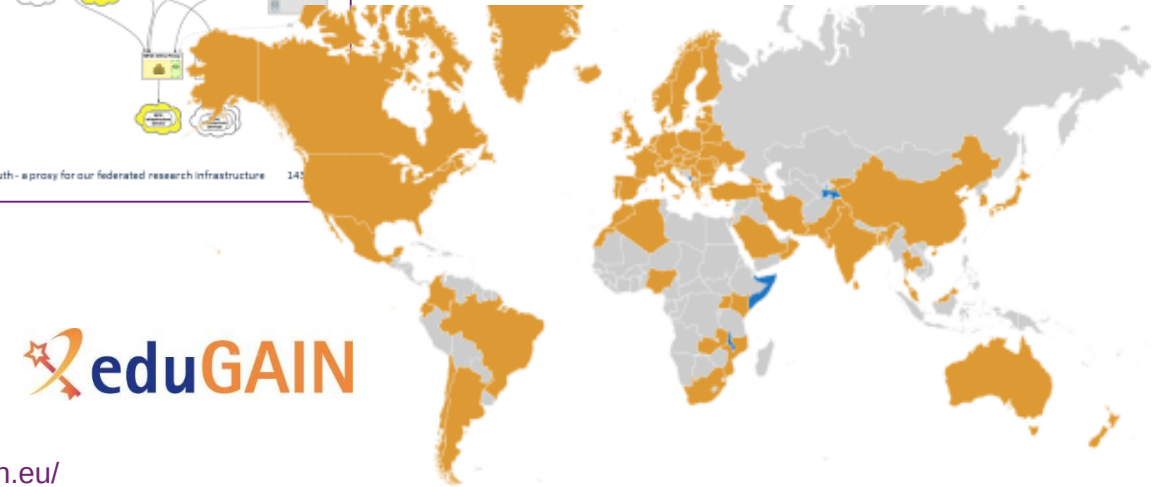
PDP-CT itself is also a collaboration!

PDP and CT-PDP – “most of whom cannot be seen”. But you can soon find out – join the Office Hours every 1st Thursday of the month!

Zie ginds weer de stoomboot (see Santa's ship sail again)

Infrastructure for Collaboration: of AARC TREEs & Tokens





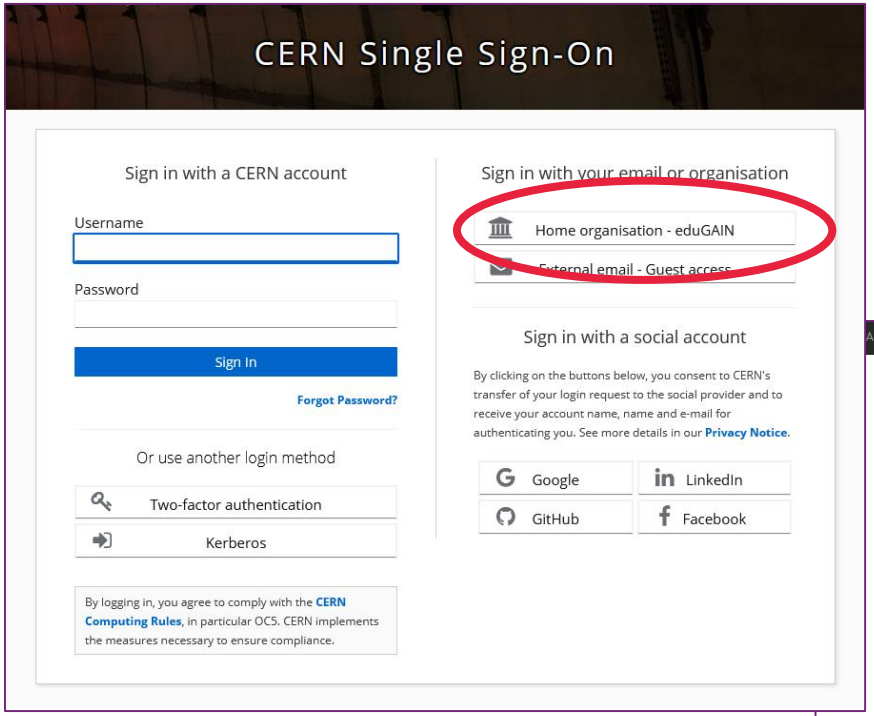
eduGAIN

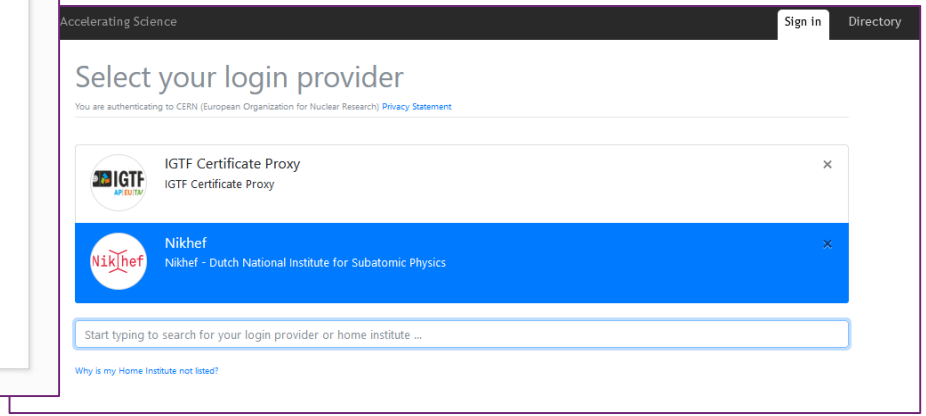
See also https://doi.org/10.2777/8702, https://rcauth.eu/

Zie ginds weer de stoomboot (see Santa's ship sail again)

Nikhef

Where could *you* use federated login? … maybe here? ☺



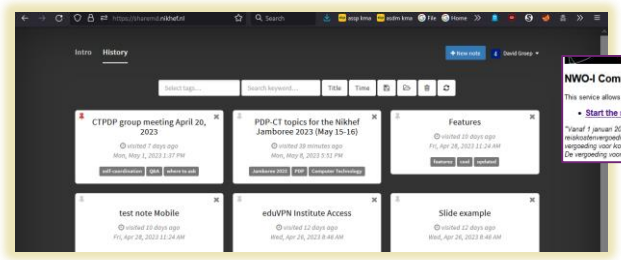


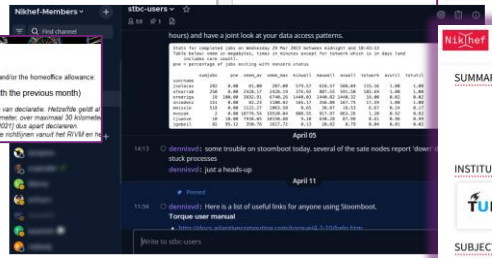
New CERN SSO architecture: Hannah Short et al. - note that this applies only to the new CERN login system, not the legacy AD one

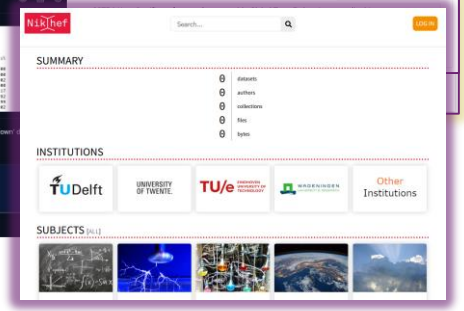
Zie ginds weer de stoomboot (see Santa's ship sail again)

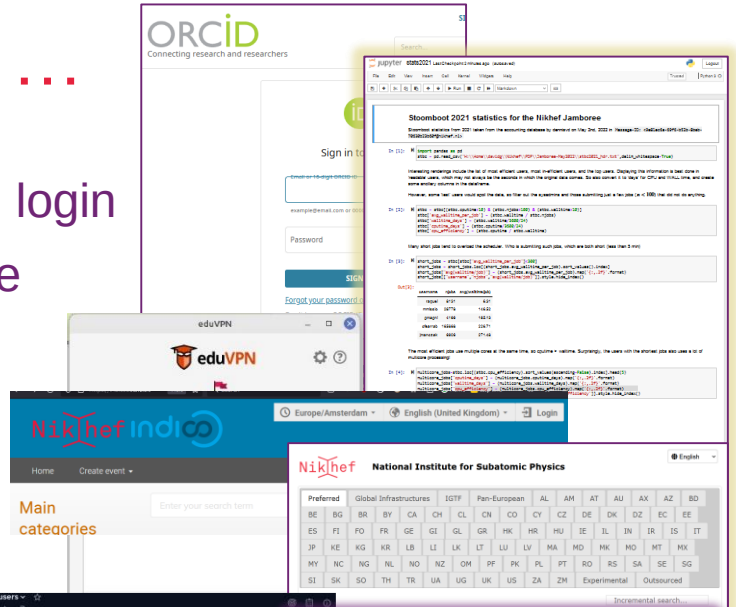
Some services you may already use …

- **Nikhef (and CERN Indico):** with global federated login
- **eduVPN:** securely access Callysto and your home
- **eVA**, **SURFdrive**, and **FileSender** to collaborate
- **Callysto:** JupyterHub with $HOME and SSO
- ***Experimental services***: ShareMD, commute, …









But do read https://www.nikhef.nl/pdp/doc/experimental-services before using experimental services ...

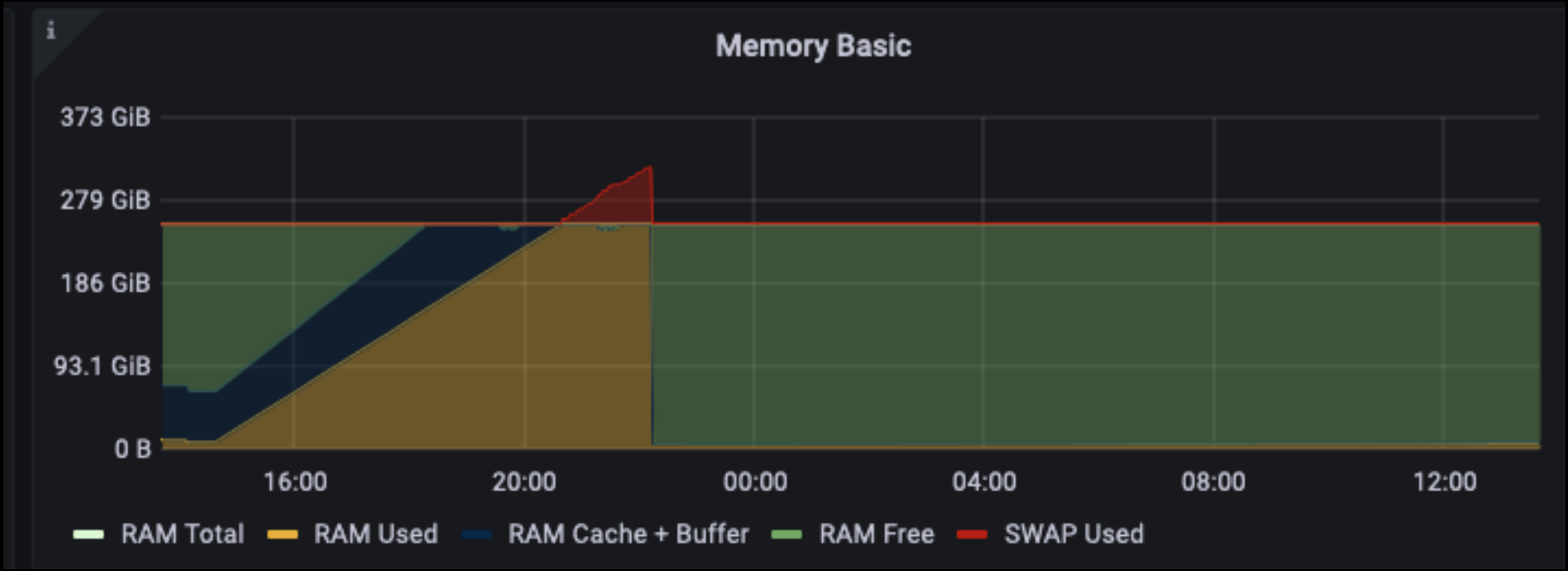
Zie ginds weer de stoomboot (see Santa's ship sail again)

Accelerated computing: need4speed (and need4scale)

State-of-the-art algorithms scale quadratically or worse with the number of signals in a given tracking detector. Time information is therefore needed to keep computational cost of pattern recognition algorithms under control while meeting required physics performance goals (tracking, vertexing, pileup suppression and particle identification)

- co-design of detectors, readout and processing is needed to optimize performance and keep costs under control.
- to provide sufficient computational resources at reasonable cost accelarators - GPUs, FPGAs, other future architectures - must be investigated to know sufficiently early which solutions are feasible.
- holds for both real-time processing, offline processing and simulation.
- existing large-scale distributed compute infrastructure - the GRID - must also adapt.

On data structures and scalability

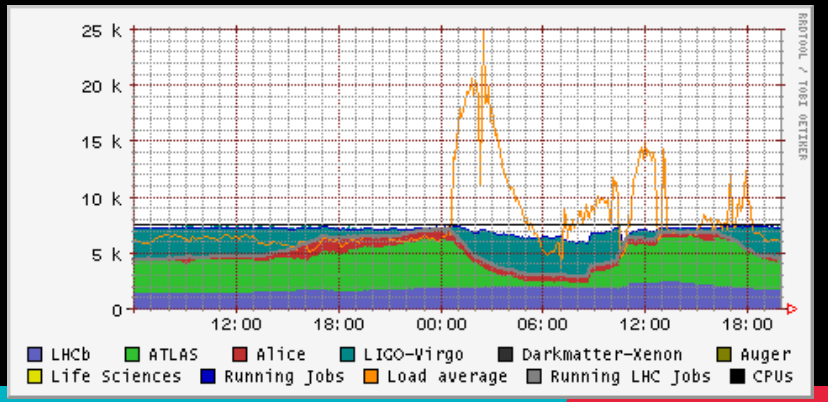


Zie ginds weer de stoomboot (see Santa's ship sail again)

Need4Scale – with great power …

```
[root@wn-pep-002 ~]# top
top - 09:40:47 up 71 days, 12:17,  2 users,  load average: 110.38, 101.43, 106.3
Tasks: 700 total,   7 running, 666 sleeping,   0 stopped,  27 zombie
%Cpu(s): 17.0 us,  2.0 sy,  0.0 ni, 81.0 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
KiB Mem : 39462902+total, 23514457+free, 10406320 used, 14907812+buff/cache
KiB Swap: 67108860 total, 66841340 free,   267520 used. 37964784+avail Mem

   PID USER      PR  NI    VIRT    RES    SHR S  %CPU %MEM     TIME+ COMMAND
 82661 expt000   20   0 5618756 396356    924 R 360.0  0.1   5:14.43 mksquashfs
 72615 expt000   20   0 5626336 248516    816 R  90.0  0.1   5:44.11 mksquashfs
 83257 expt000   20   0 5611608 219300    852 S  90.0  0.1   1:17.66 mksquashfs
 ...
```

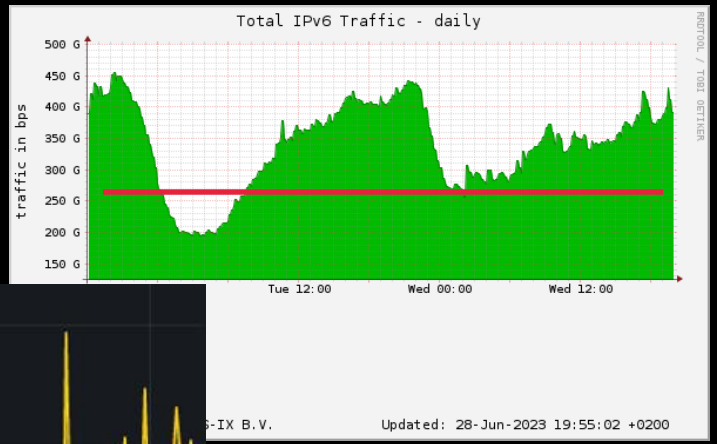
User doing mass *creation* of containers, rebuilding their python ‘virtual env’ for each job, running on >> 4000 cores



June 28th, 2023, data from Nikhef NDPF stats & cricket (top),

Zie ginds weer de stoomboot (see Santa's ship sail again)

… comes great responsibility!



1/4th of all IPv6 traffic that night at the Amsterdam Internet Exchange was from this one user @NDPF

Pulling the python packages at line rate and downloading public python repositories ultimately *will* flood SURFnet (and suck up Cloudfare's IPv6)

Zie ginds weer de stoomboot (see Santa's ship sail again)

June 28th, 2023,
asd001b-jnx-01 to asd001b-jnx-04 (left),
AMS-IX SFlow https://stats.ams-ix.net/sflow/index.html (bottom)

Other ways to ‘exercise’ a cluster: jobs of which life came to an unfortunate premature end

Actually this isn't the right metric though, as there could be scads of short jobs with a few very long jobs mixed in, reaching a reasonable average duration. Let's look at the median instead.

The query:

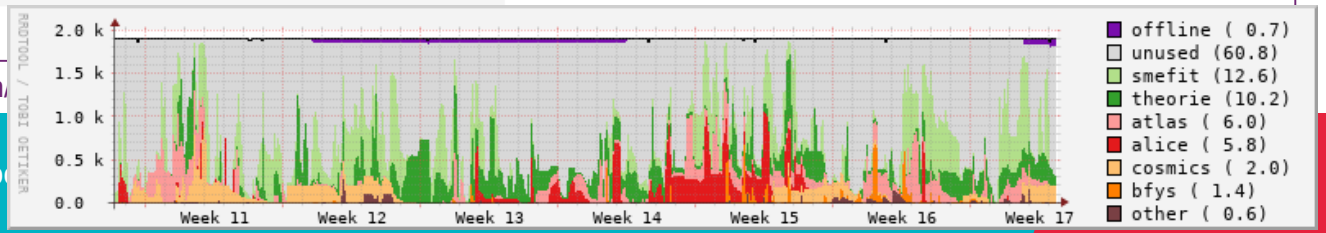
```
$ cat stoomboot-shortjobs
SELECT
        userid.id,
        median(job.WallTi
        sum(job.CPUTime)/
        count(*) as njobs
FROM
        job, userid
WHERE
        job.userid=useric
        job.facility=1 AN
        job.EndTime > '20
        job.EndTime < '20
GROUP BY userid.id HAVING
ORDER BY medianrun ASC;
```

```
In [7]: dfshort_2024=pd.read_table('shortjobs-2024.tsv', index_col='id')

        # we recommend minimum job duration of 60 seconds;
        # show rows for which the mean run time is less than that

        dfshort_2024[dfshort_2024.medianrun<60.0]
```

Out[7]:

id	medianrun	cpsum	njobs
msaharan	0.0	0.8569	162301
jlomker	0.0	4.1052	51228

https://www.nikhef.nl/pdp/stats/stbc/intern/



Zie ginds weer de stoombo

And this year, the stoomboot winner is …

Congrats to Tanishq (theory), Mohit (cosmics), Efren (LHCb), and Giacomo (theory)

```
In [3]:
lastyear=df2023.add(df2024, fill_value=0)
lastyear['njobs'] = lastyear['njobs'].astype(int)

In [4]:
sly = lastyear.sort_values('cpsum', axis=0, ascending=False)
```

Users ranked by actual computing work

Out[4]:

id	corewallsum [years]	wallsum [years]	cpsum [years]	njobs
tsharma	62.4690	53.8772	58.2098	169826
msaharan	47.9864	47.9864	47.5289	232255
erodrigu	39.9191	39.9191	36.7769	185849
gmagni	44.7394	3.7516	28.8070	87206
				45721
				19970
				291918
pkrack	23.5386	6.8287	20.2963	42307
pveen	19.7364	19.7364	19.6241	47018
bkortman	20.9752	20.9752	18.9141	701677

Stoomboot usage May 2023 – now, analysis: Jeff Templon
See also https://www.nikhef.nl/pdp/stats/stbc/intern/stbc_summ_plot

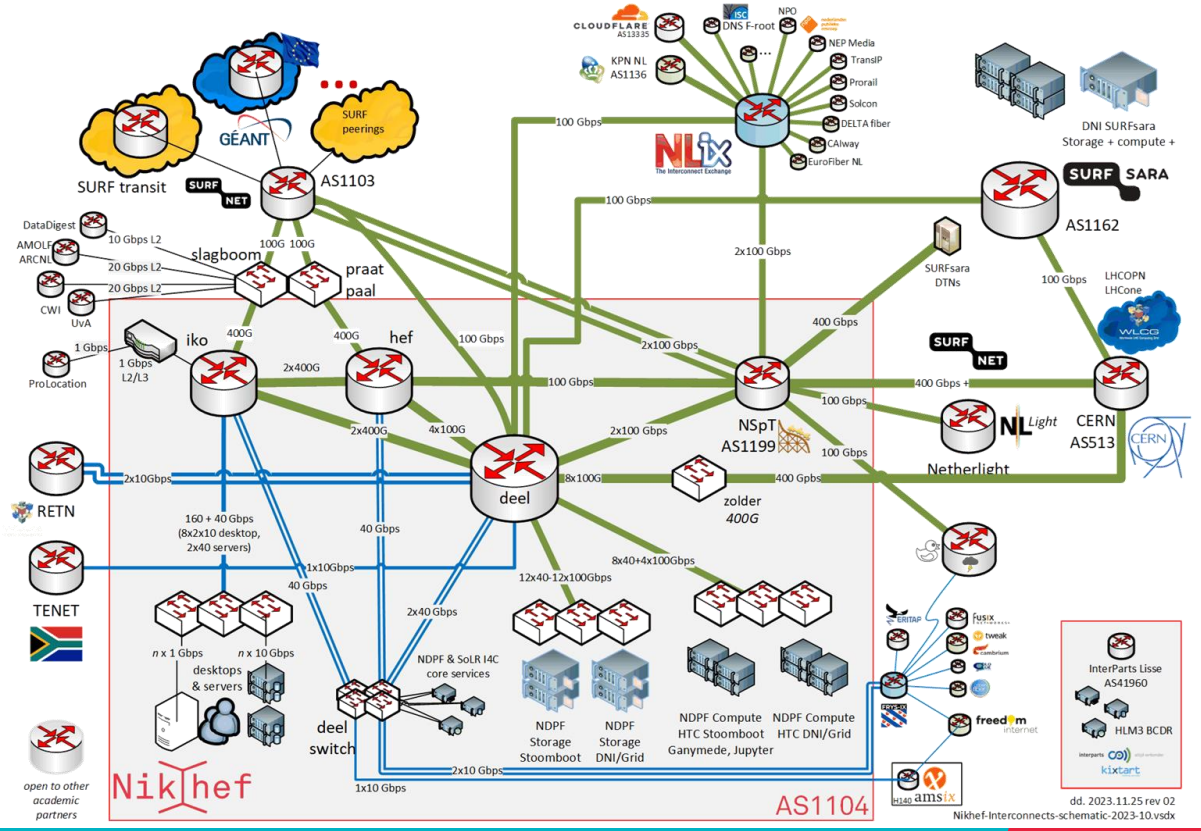
Zie ginds weer de stoomboot (see Santa's ship sail a

Computing can make an image with many tracks as well ☺

Nikhef AS1104
internal network
and private peerings

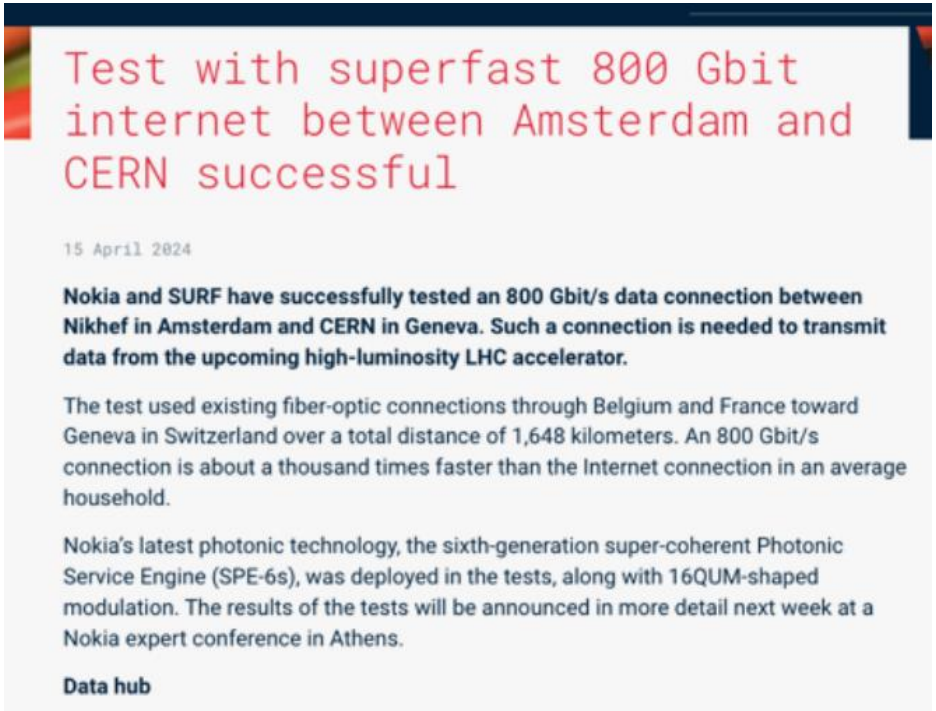
Time for some track finding!
Curious about your way home?
From home, use

```
traceroute -A myip.nikhef.nl
```



Zie ginds weer de stoomboot (see Santa's ship sail again)

In time for the HL HLC (and more)



Test with superfast 800 Gbit internet between Amsterdam and CERN successful

15 April 2024

Nokia and SURF have successfully tested an 800 Gbit/s data connection between Nikhef in Amsterdam and CERN in Geneva. Such a connection is needed to transmit data from the upcoming high-luminosity LHC accelerator.

The test used existing fiber-optic connections through Belgium and France toward Geneva in Switzerland over a total distance of 1,648 kilometers. An 800 Gbit/s connection is about a thousand times faster than the Internet connection in an average household.

Nokia's latest photonic technology, the sixth-generation super-coherent Photonic Service Engine (SPE-6s), was deployed in the tests, along with 16QUM-shaped modulation. The results of the tests will be announced in more detail next week at a Nokia expert conference in Athens.

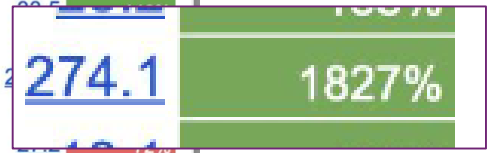
Data hub

Zie ginds weer de stoomboot (see Santa's ship sail again)

ATLAS T0 export repeat

- T0 export rates are the most important use case and were not achieved
- The rates weren't achieved because they were queued behind production
  - T2 traffic is non negligible in ATLAS (42% dst, 25% src)
- Tests were repeated injecting one site at the time
  - Rates improved for the majority of sites
- Some differences:
  - SARA was testing 800 Gb/s after DC24; was injected with much larger rates
  - RAL wanted to test writing directly to tape in the second test; other limitations were identified
  - NDGF resolved the dcache bug that was affecting them

Site	T0 Export	DC24 best rates on day 1,2	% of expected rates	T0-T1 one T1at the time	% of expected rates
BNL-ATLAS	60	31.5	53%	61.3	102%
FZK-LCG2	32	26.4	83%	42.2	132%
IN2P3-CC	38	43	113%	50.9	134%
INFN-T1	23	19.3	84%		
NDGF-T1	15	13.8	92%		
SARA-MATRIX	15	12.2	81%		
pic	11	12.3	112%		
RAL-LCG2	38	15	39%		
TRIUMF-LCG2	25	23.9	96%	27.2	109%
T1 summary	257	197.4	77%	562.7	219%
T1 summary -SARA	242	185.2	77%	288.6	119%



T1 summary	257	197.4	77%	562.7	219%
T1 summary -SARA	242	185.2	77%	288.6	119%

LHCONE/LHCOPN meeting, April 2024

25

800Gbps in practice

From: Katy Ellis (STFC RAL): DC24 report, LHCOPN meeting, Catania, IT, April 2024 https://indico.cern.ch/event/1349135/#6-dc24-report

Zie ginds weer de stoomboot (see Santa's ship sail again)

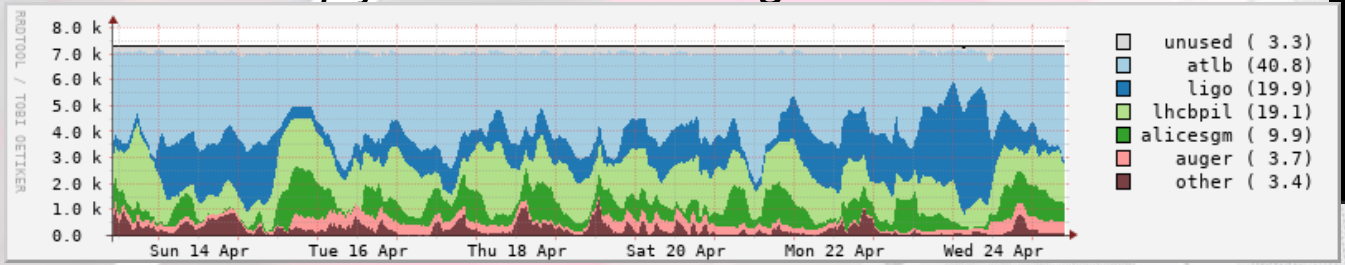
Nikhef

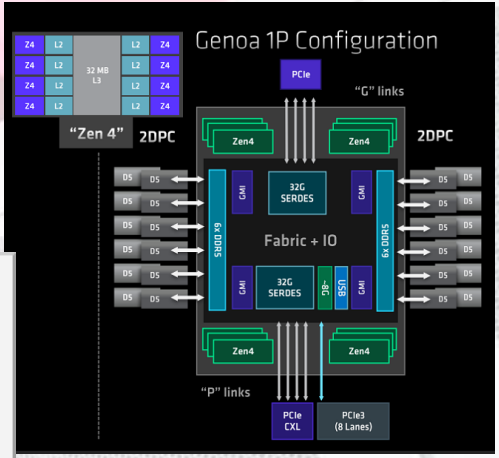
… towards ~ tomorrow

But why not get results faster on our next-gen clusters?

- today's Genoa already +30% HEPscore performance
- new clusters ordered the end of 2023 …

and this 'vuurpijl' extension being installed now …





and 'containerise' your work in the future!

- better access to **GPUs on stoomboot**
- run on our **newer hardware** *that will use* Alma 9 & Debian
- prepare for **data analysis preservation** and good research data management

Screenshots: AMD EPYC Genoa architecture; archive.nikhef.nl "HADRON" RDM service

Zie ginds weer de stoomboot (see Santa's ship sail again)

Next up

“Scavenging Containers fallen off the Stoomboot – how Condor & containers make analysis flexible” (Dennis van Dok)

David Groep
davidg@nikhef.nl
https://www.nikhef.nl/~davidg/presentations/
https://orcid.org/0000-0003-1026-6606

Nikhef