SURF

LUMI

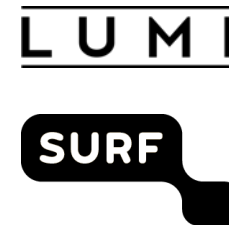## LUMI in The Netherlands

Henk Dreuning (SURF & LUST)

May 3, 2024

Netherlands and the LUMI consortium

- The Netherlands joined the LUMI consortium on March 19, 2024
 - Own share of LUMI resources will be available to Dutch researchers

- Envelope size 2024:

CPU core-hours	GPU-hours	TB-hours
22 million	1.4 million	10.3 million

- LUMI serves as an extension of our national compute facilities (Snellius)
  - Serves growing demand for compute resources
  - Can accommodate extreme-scale projects

- You will be able to request access to the Dutch share in June (expected)

- SURF will facilitate access to LUMI
  - Applications for compute time supervised in collaboration with NWO

- Two types of grants: LUMI pilot and LUMI regular
  - Similar to the small/large compute applications for the national system Snellius

Snellius and LUMI
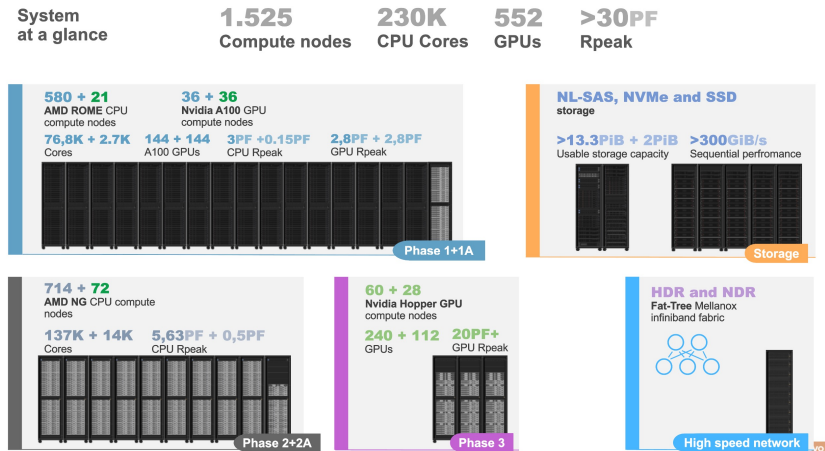
- Snellius
    - 1387 CPU nodes (AMD Rome and Genoa)
    - 72 GPU nodes (160 after phase 3) with NVIDIA GPUs
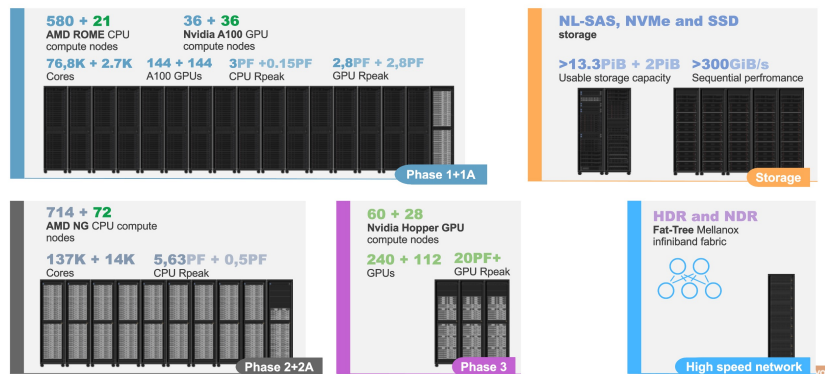   - Infiniband network, Fat-tree topology

System at a glance — 1.525 Compute nodes, 230K CPU Cores, 552 GPUs, >30PF Rpeak

580 + 21 AMD ROME CPU compute nodes — 76,8K + 2.7K Cores, 144 + 144 A100 GPUs, 3PF +0.15PF CPU Rpeak, 2,8PF + 2,8PF GPU Rpeak

36 + 36 Nvidia A100 GPU compute nodes

NL-SAS, NVMe and SSD storage — >13.3PiB + 2PiB Usable storage capacity, >300GiB/s Sequential perfromance — Storage

714 + 72 AMD NG CPU compute nodes — 137K + 14K Cores, 5,63PF + 0,5PF CPU Rpeak — Phase 2+2A

60 + 28 Nvidia Hopper GPU compute nodes — 240 + 112 GPUs, 20PF+ GPU Rpeak — Phase 3

HDR and NDR Fat-Tree Mellanox infiniband fabric — High speed network

Phase 1+1A

Snellius and LUMI

- Snellius
  - 1387 CPU nodes (AMD Rome and Genoa)
  - 72 GPU nodes (160 after phase 3) with NVIDIA GPUs
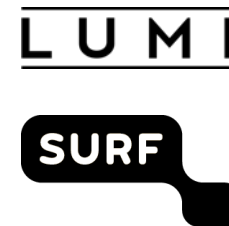 - Infiniband network, Fat-tree topology

- LUMI
  - 2048 CPU nodes (AMD Milan)
  - 2978 GPU nodes with AMD GPUs
 - HPE Cray Slingshot-11 interconnect, Dragonfly topology
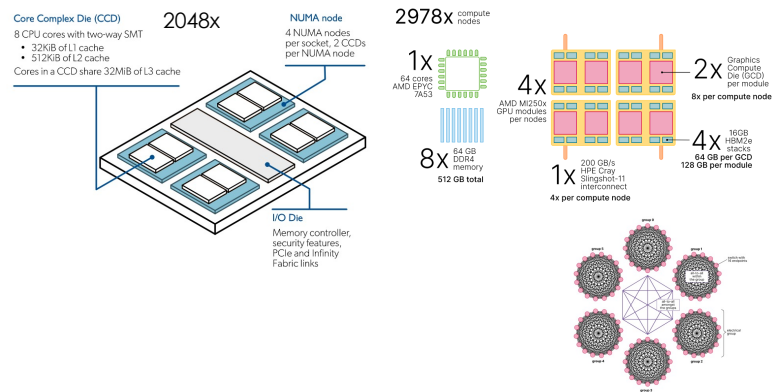
Which projects can apply for LUMI access?

- Projects that require LUMI's scale
  - Large-scale GPU projects
    - E.g. distributed training of deep neural networks
    - More than ~32-64 GPUs

- Projects interested in LUMI specifics
  - AMD hardware and tools
  - HPE Cray software stack and tools
  - Slingshot interconnect and Dragonfly network architecture

- For which projects is Snellius the better option?
  - Large-scale CPU-only projects
  - Projects that require NVIDIA GPUs

LUMI Pilots

- For projects requesting up to
  - 500.000 CPU core-hours
  - 14.000 GPU hours
  - 100.000 TB-hours

- Project duration up to 1 year

- Technical review only, no scientific review
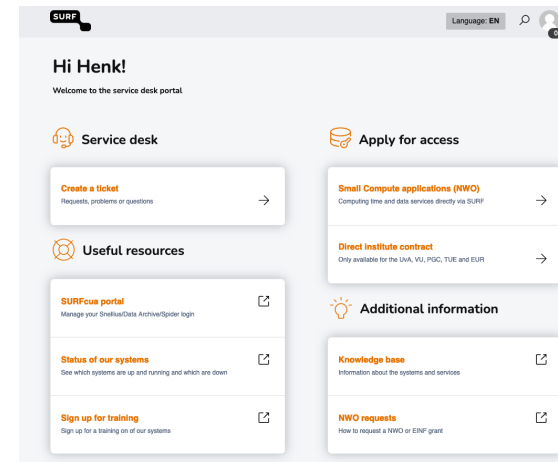 - Processing time: 1-2 weeks
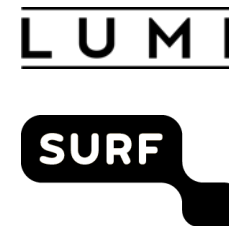
LUMI Regular
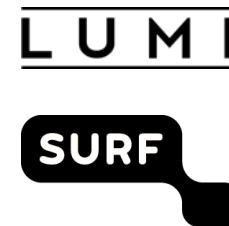
- For projects requesting more than
  - 500.000 CPU core-hours
  - 14.000 GPU hours
  - 100.000 TB-hours

- Project duration up to 1 year

- Scientific and technical review
  - Processing time: max 4 months / WGS meets every 2 months
  - Preliminary access within 1 month

- Requirements
  - You had prior access to LUMI through a LUMI pilot or EuroHPC call
  - You can demonstrate that your code is ready for LUMI: it runs and scales well

How to apply

- LUMI pilots: directly with SURF, through E-Infra portal
  - You can simultaneously apply for other SURF services too

How to apply

- LUMI pilots: directly with SURF, through E-Infra portal
  - You can simultaneously apply for other SURF services too

Resources
- [ ] Snellius
- [ ] Research Cloud - HPC Cloud
- [ ] Data processing - Grid
- [ ] Data processing - Spider
- [ ] Cloud Research Consultancy - MS4
- [ ] Cloud Research Consultancy - SDA
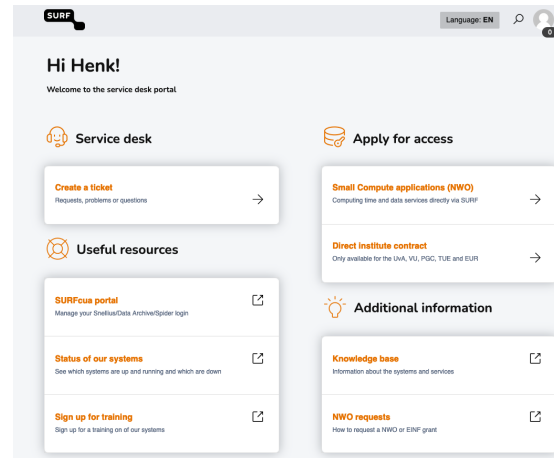- [ ] I am not sure. Please contact me.

How to apply

- LUMI pilots: directly with SURF, through E-Infra portal
  - You can simultaneously apply for other SURF services too

Resources
- ☐ Snellius
- ☐ Research Cloud - HPC Cloud
- ☐ Data processing - Grid
- ☐ Data processing - Spider
- ☐ Cloud Research Consultancy - MS4
- ☐ Cloud Research Consultancy - SDA
- ☐ I am not sure. Please contact me.

- LUMI regular: with NWO, through ISAAC portal
  - LUMI integrated in application form for national compute services

Hi Henk!
Welcome to the service desk portal

Service desk

Create a ticket
Requests, problems or questions

Useful resources

SURFcua portal
Manage your Snellius/Data Archive/Spider login

Status of our systems
See which systems are up and running and which are down

Sign up for training
Sign up for a training on of our systems

Apply for access

Small Compute applications (NWO)
Computing time and data services directly via SURF

Direct institute contract
Only available for the UvA, VU, PGC, TUE and EUR

Additional information

Knowledge base
Information about the systems and services

NWO requests
How to request a NWO or EINF grant

Rekentijd nationale computersystemen

Via de call for proposals Rekentijd Nationale Computersystemen 2024 kunnen onderzoekers en onderzoeksgroepen toegang kunnen krijgen tot rekentijd en bijbehorende datadiensten en expertise op de nationale geavanceerde computersystemen Supercomputer Snellius, Data Processing (Grid/Spider), Cloud Research Consultancy en HPC Cloud (via SURF Research Cloud).

① Oriënteren	② Voorbereiden	③ Indienen

Waarvoor ⊕

Voor wie ⊕

Wat aanvragen ⊕

Wanneer ⊕

Beoordeling ⊕

Meer informatie ⊕

EuroHPC calls

- LUMI access is also possible through EuroHPC calls

Access mode	CPU core-hours	GPU-hours	TB-hours	Cut-off frequency	Max time-to-resources-access	Project duration
Benchmark	<= 256K	<= 10K	<= 65K	Monthly	2-3 weeks	3 months
Development	<= 512K	<= 18K	<= 130K	Monthly	2-3 weeks	1 year (renewable)
Regular	7.68M - 14.59M	140K - 630K	Unspecified	Twice per year	4 months after cutoff	1 year
Extreme	> 15.36M	> 700K	Unspecified	Twice per year	6 months after cutoff	1 year

- Advantages of applying for the Dutch share:
  - Shorter time-to-access (review process)
  - Familiar application process
  - Access to SURFs L3 support

SURF support for LUMI
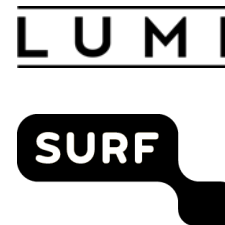
- L1 and L2 support through LUST

- L3 support through SURF's 'promising application' program

- What can SURF support you with?
    - Porting your code to LUMI (AMD GPUs)
    - Benchmarking and profiling your code
    - Optimizing your code's performance
    - Improving your code's scalability

- How to apply?
    - Consultancy hours requested during compute time application
        - Up to 160 hours for LUMI pilots and 520 for LUMI regular projects
    - You will be contacted by a SURF advisor
    - We will assess whether we can provide requested support

SURF

LUMI

Questions?