Hoe kunnen we AI-faciliteiten duurzaam exploiteren?

Huidige inzichten en toekomstige ontwikkelingen

Axel Berg

18 mei 2026, ECP Deelnemersspecial - Investeren in digitale infrastructuur

We moeten AI-faciliteiten duurzaam exploiteren

Huidige inzichten en toekomstige ontwikkelingen

Axel Berg

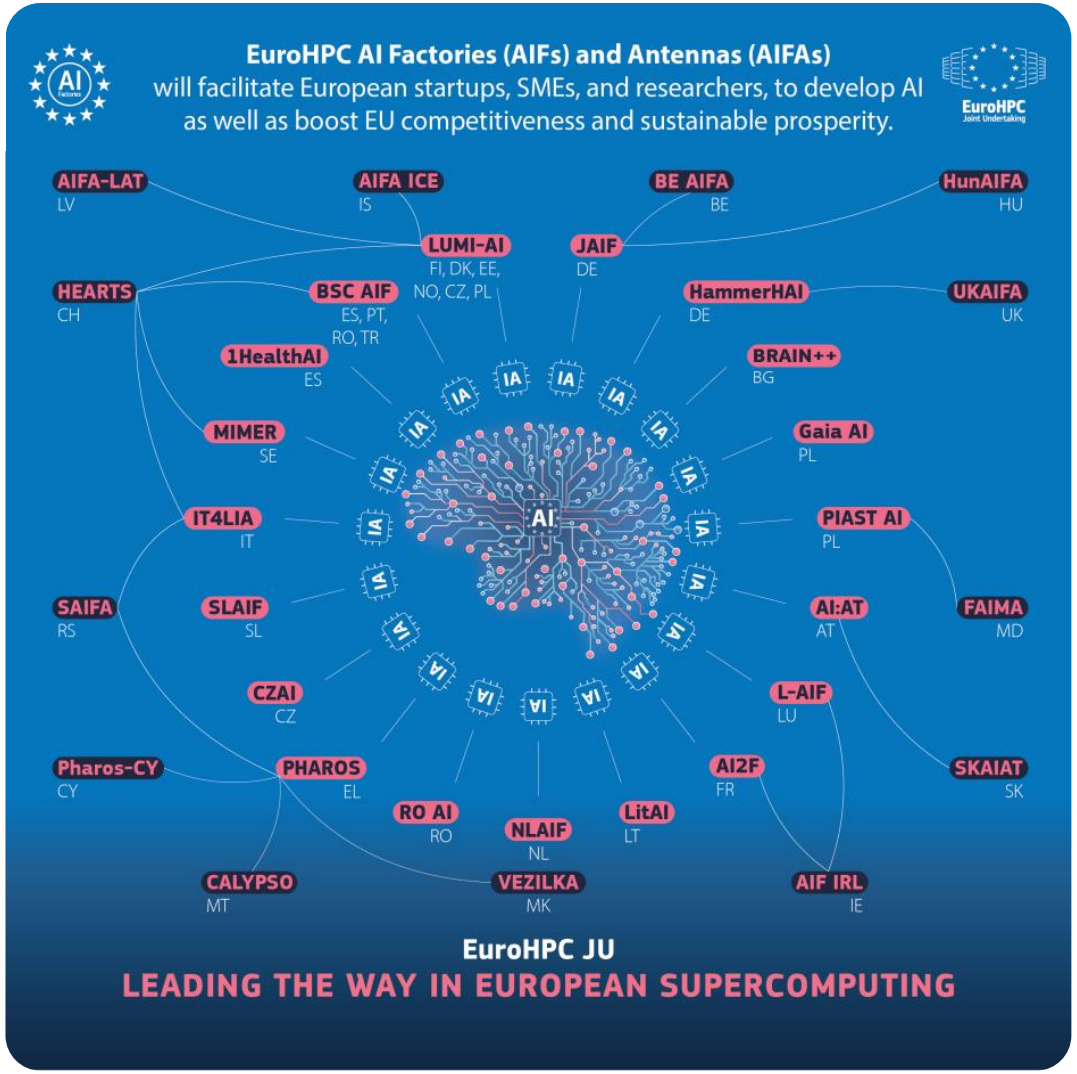
18 mei 2026, ECP Deelnemersspecial - Investeren in digitale infrastructuur

Snellius - Nationale supercomputer

- 40+ jaar nationale HPC dienstverlening
- > 240.000 cores
- Energieverbruik: ca 1.5MW
- Gebruik voor wetenschappelijk onderzoek

Nederlandse AI Fabriek

- **Creëren van economische, maatschappelijke en wetenschappelijke impact**
- Grote AI-supercomputer + Expertise centrum
- In provincie Groningen
- AI-Facility Operationeel eind 2027-begin 2028
- Target gebruikers:
  - MKB en start-ups
  - Onderzoek en onderwijs
  - Overheid
- Partners: AIC4NL, TNO, Samenwerking Noord, SURF
- Financiering NL + Nij Begun + EuroHPC JU

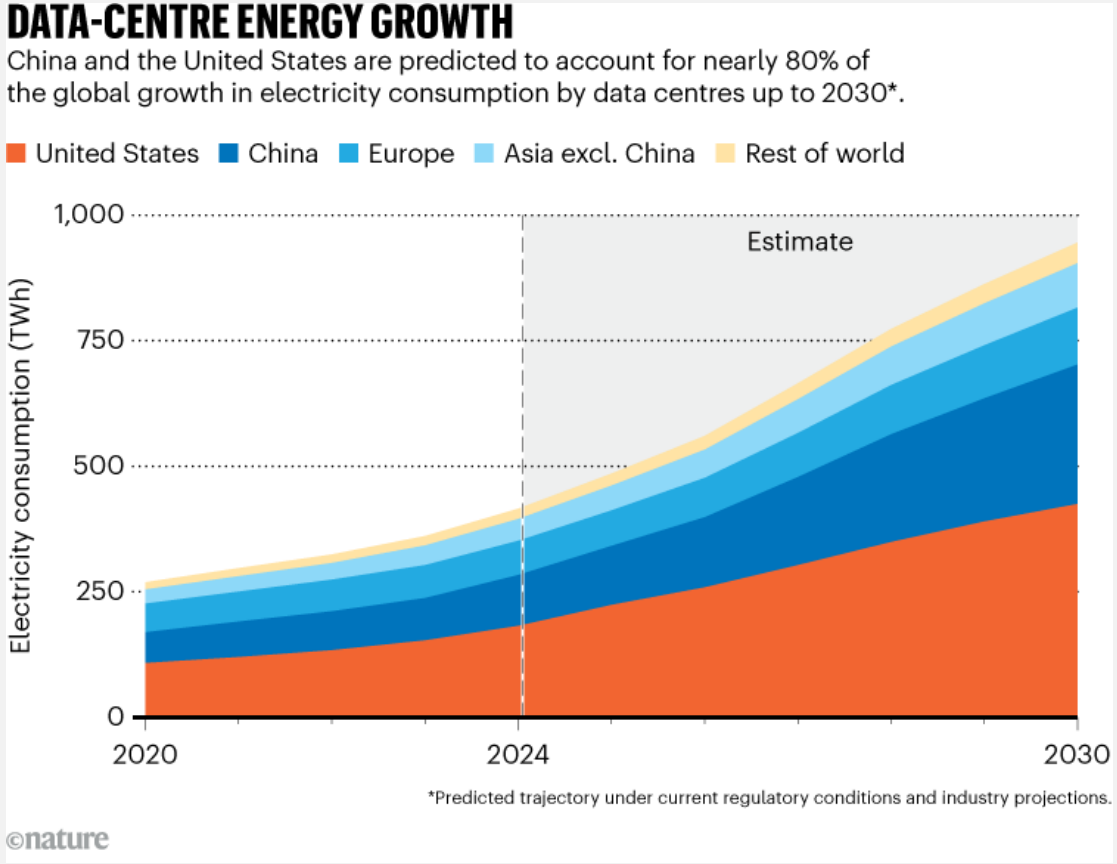


Trends zijn duidelijk

- **Datacenter energiegebruik groeit snel, primair gedreven door AI**
- Globaal energiegebruik van datacenters is nu 3-4%
- Hyperscalers bouwen Gigawatt Scale Machine Learning (**100MW** ML implementatie is **routine**)

- **Rack energiedichtheden (kW/rack) groeien enorm**

Data centres will use twice as much energy by 2030 – driven by AI



www.nature.com/articles/d41586-025-01113-z

11M cores, 35MW

400kW/rack

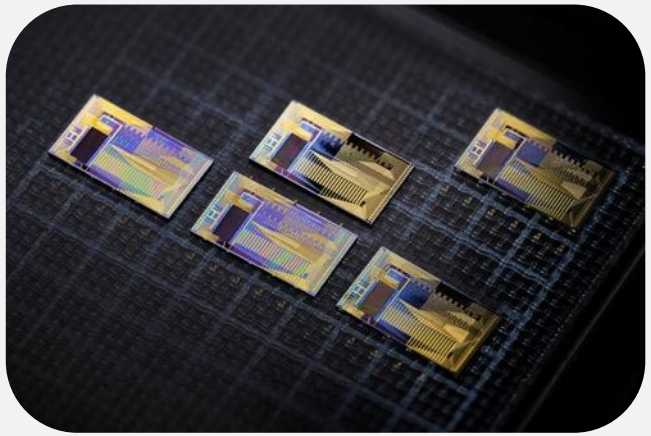
Up to 1 MW/rack

and up to 3.500+ kg/rack

Toekomst

- **Energiebesparingen**: veel kansen, maar vragen veel onderzoek, niet alleen technologie
  - Algorithmisch
  - Agentic AI voor datacenter en systeemmanagement
  - Nieuwe energie-efficiente technologieen
    - Gespecialiseerde AI chips
    - Quantum: complementair
    - Neuromorphic
    - Photonic
- **Experimenteren**: radicaal breken met traditionele aanpakken en werkwijzen

https://www.tue.nl/en/news-and-events/news-overview/14-09-2023-breakthrough-way-to-train-neuromorphic-chips



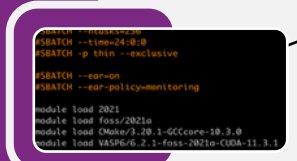
https://www.tue.nl/en/news-and-events/news-overview/22-09-2025-on-the-cusp-of-a-major-chip-transition

Op welke niveaus kunnen we acteren?

Gebruiker

Applicatie

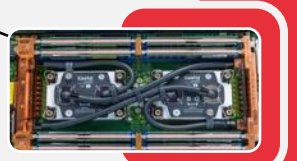
IT gebruik



Energiebron

Datacenter

IT hardware

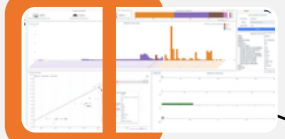


SURF

Wat kunnen we doen?

Gebruiker

Bewustwording: energiegebruik dashboard

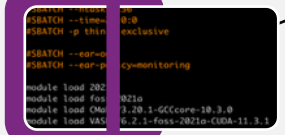


Applicatie

Efficientie: sw optimalisatie van belangrijkste applicatiies

IT gebruik

Tuning HW energie gebaseerd op running applicatie



Energiebron

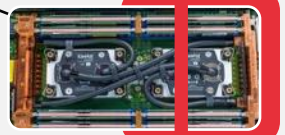
Acquisitie: % herbruikbare energie

Datacenter

Lage PUE: efficiente koeling

IT hardware

Selectie van energie-efficiente HW

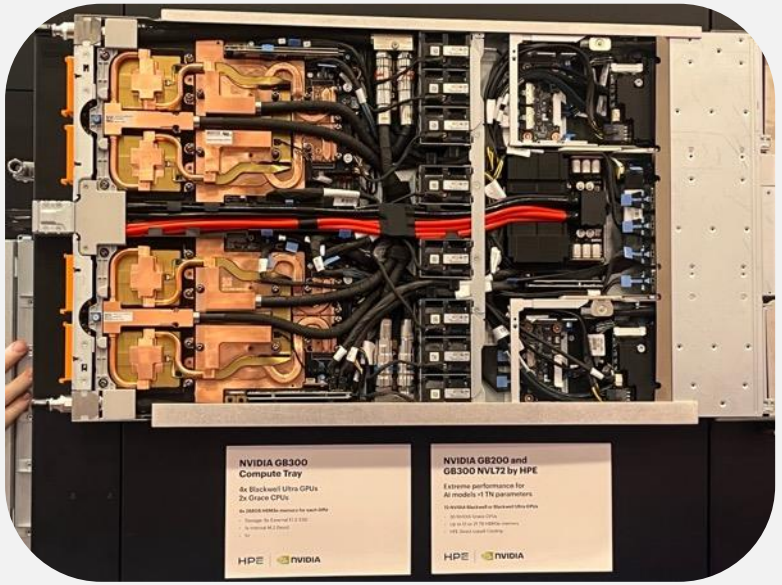


SURF

Selectie van energie-efficiente HW

- **Selecteer beste** Performance/Watt **chips**
  - General purpose en specialized
- **Gebruik TCO** (CAPEX + energiekosten) voor HW aanschaf

- **Direct Liquid Cooling** (DLC)
  - Using cold plates
  - Geen luchtkoeling
  - Geen fans
- Rear-door heat exchangers als DLC niet mogelijk is



Efficiente datacenter koeling

- Selecteer DLC servers met **warm water koeling**: inlaat 36-40ºC, uitlaat 45-55ºC
- → Bijna **100% vrije koeling** → geen energie voor koeling → **lage PUE (< 1.15)**
- Dry chillers, geen watergebruik voor koeling

- Hoge outlet temp. is voordelig voor **hergebruik van warmte** (locatie afhankelijk)

- Electriciteit redundantie alleen voor essentiele systeemcomponenten

https://www.datacenterknowledge.com/cooling/could-water-free-data-centers-move-from-concept-to-reality-

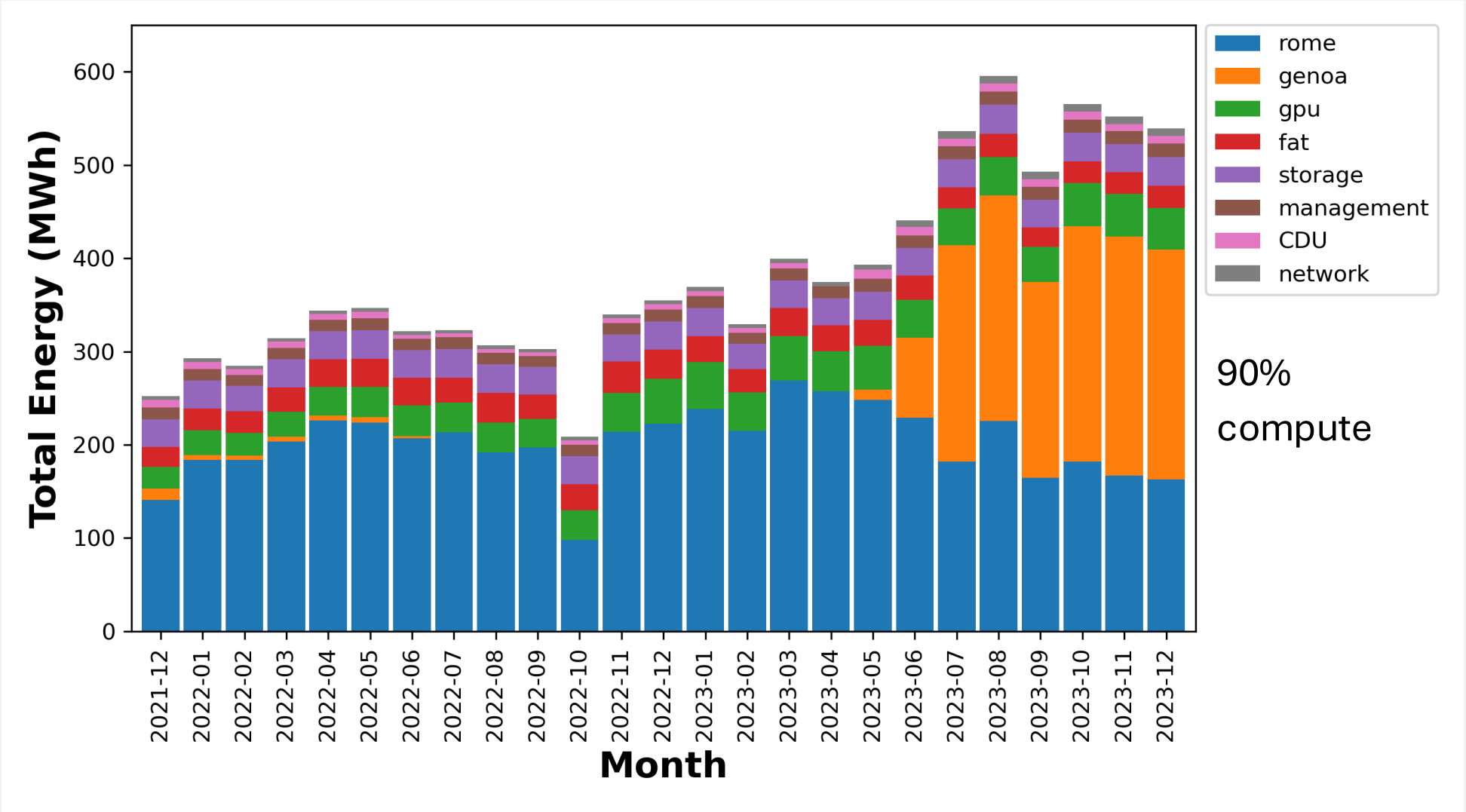
Energie system optimalisatie

Application Monitoring
Performance
Power usage

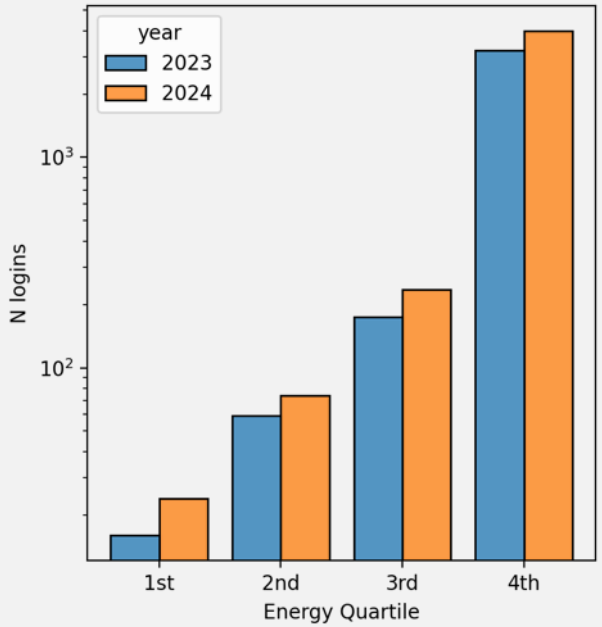
Dynamic Optimization of Applications

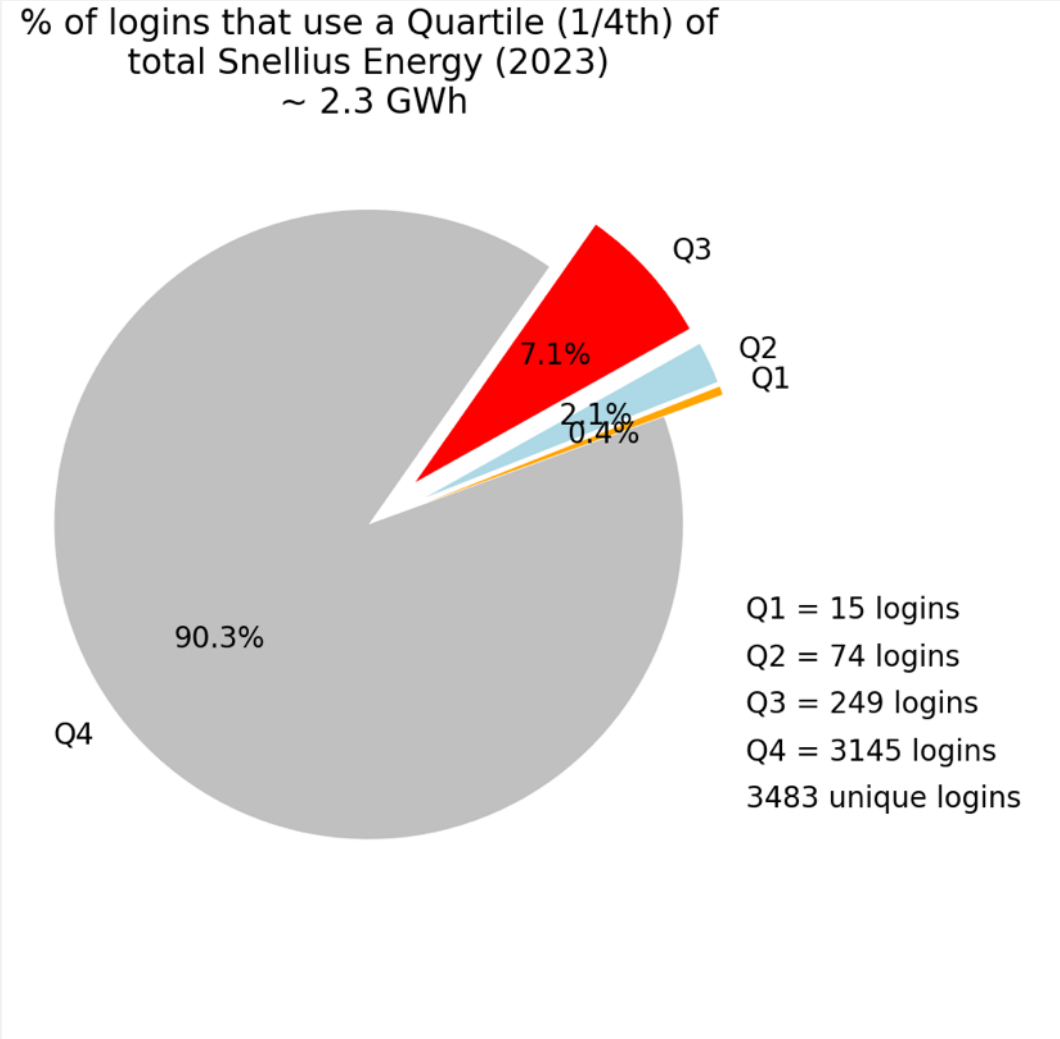
System Monitoring

Snellius - Totaal Energie Gebruik over Systeem Componenten



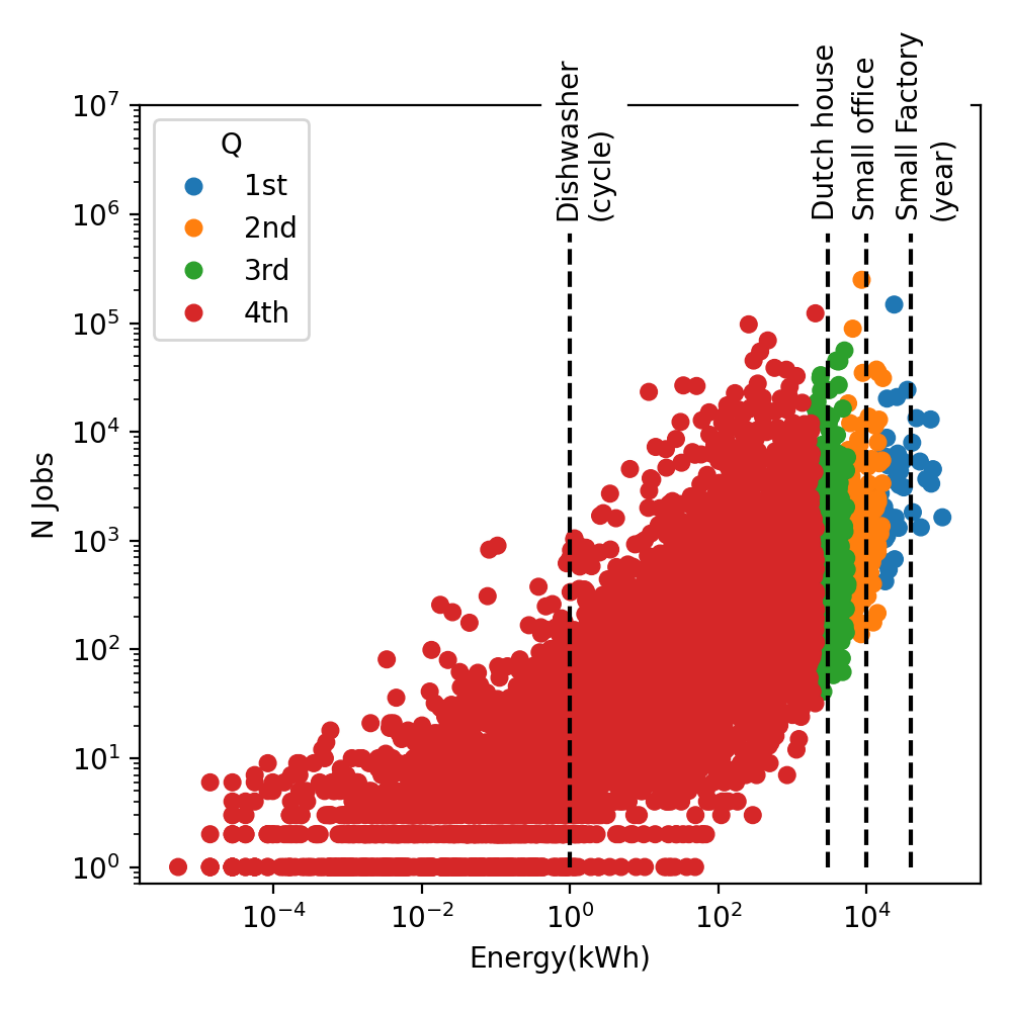
Snellius – Energie Breakdown per Gebruiker





- Slechts 0.4 % van de gebruikers (15 logins) gebruikt 25% van de energie
- 10% van de gebruikers gebruikt 75% van de energie

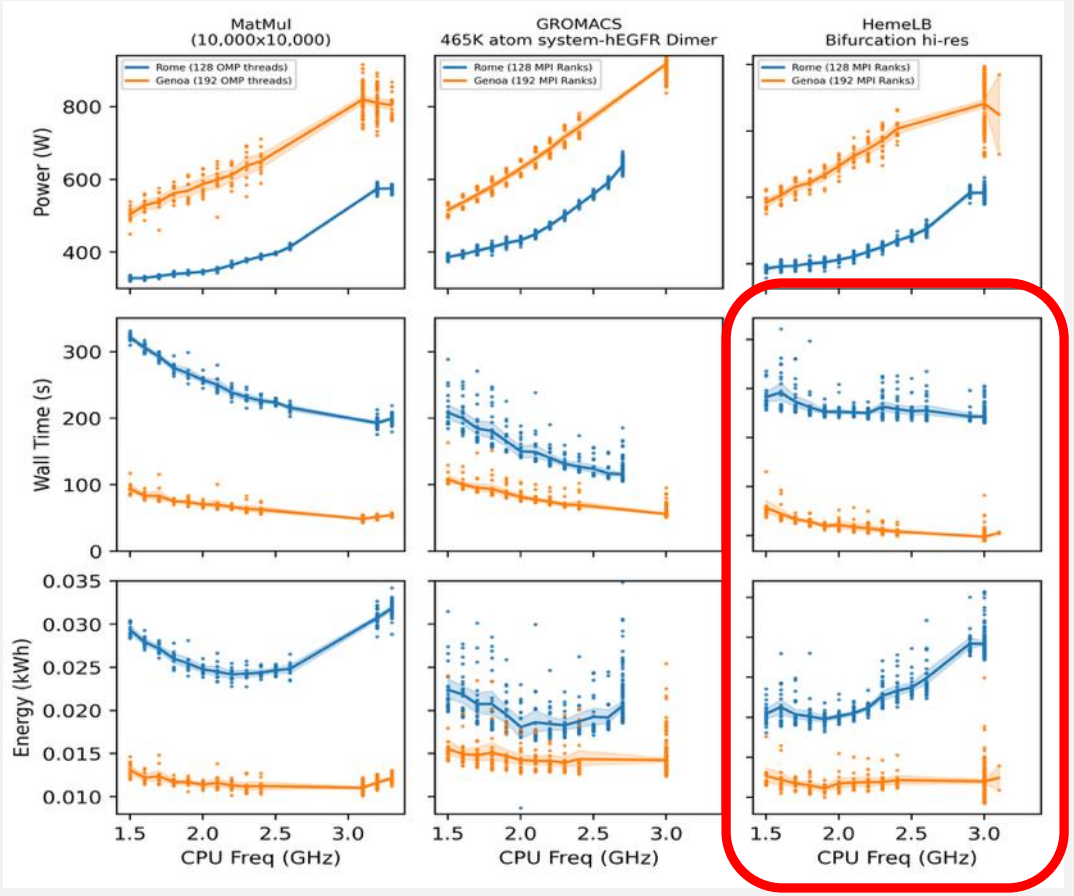
Snellius – Energie Breakdown per Gebruikers



- De meeste gebruikers verbruiken ongeveer evenveel energie als één vaatwasprogramma
- De grootste gebruikers (met energieverslindende toepassingen) zijn vergelijkbaar met een kleine plaatwerkfabriek (per jaar)

Energy Aware Runtime (EAR)

- Energy management framework voor supercomputers
- Applicatie energie metrics monitoring en tuning
- ‘Dynamic Voltage Frequency Scaling’: selectie van CPU/GPU frequenties gebaseerd op het applicatiegedrag ‘on the fly’
- **Minimize energy to solution**: reduceert CPU/GPU frequentie om energie te bespraren met een maximum tijdstraf



https://www.eas4dc.com/

Energy Aware Runtime (EAR)

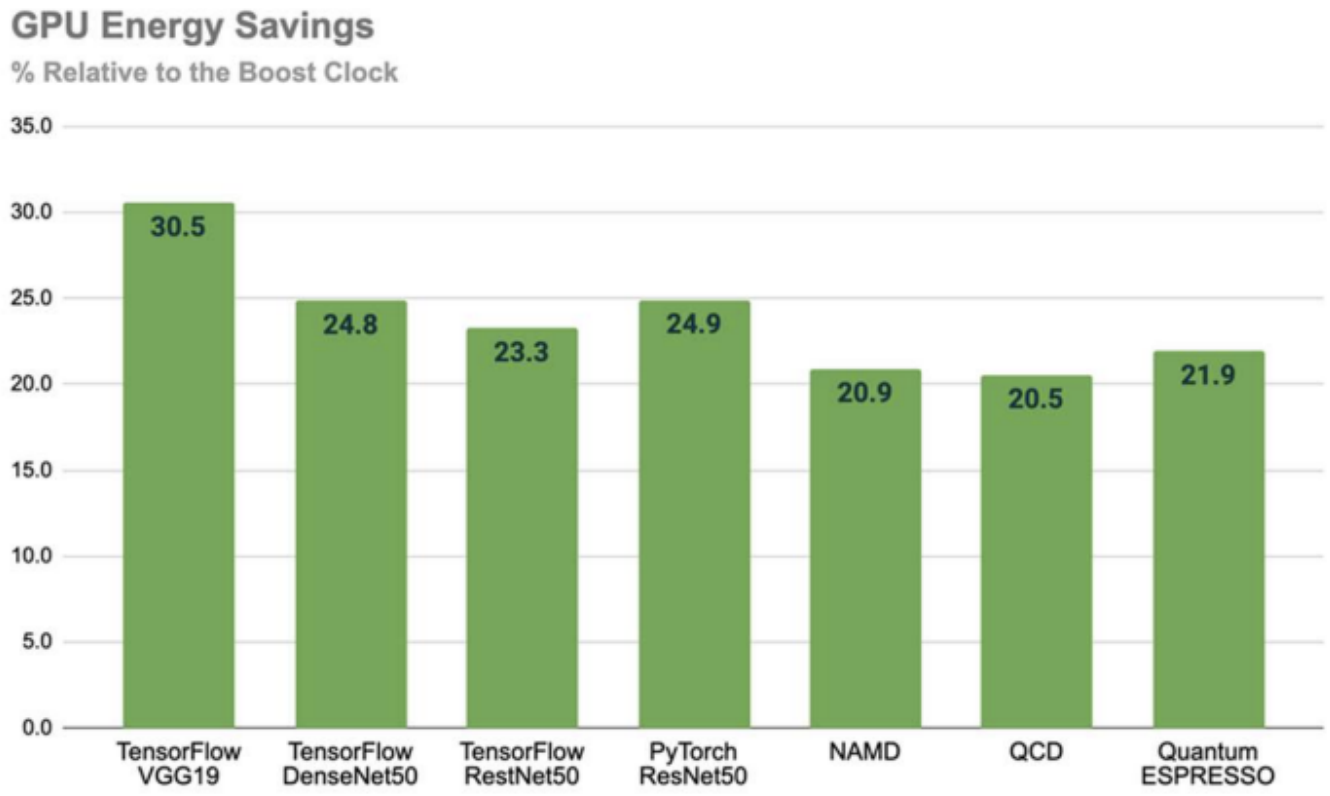
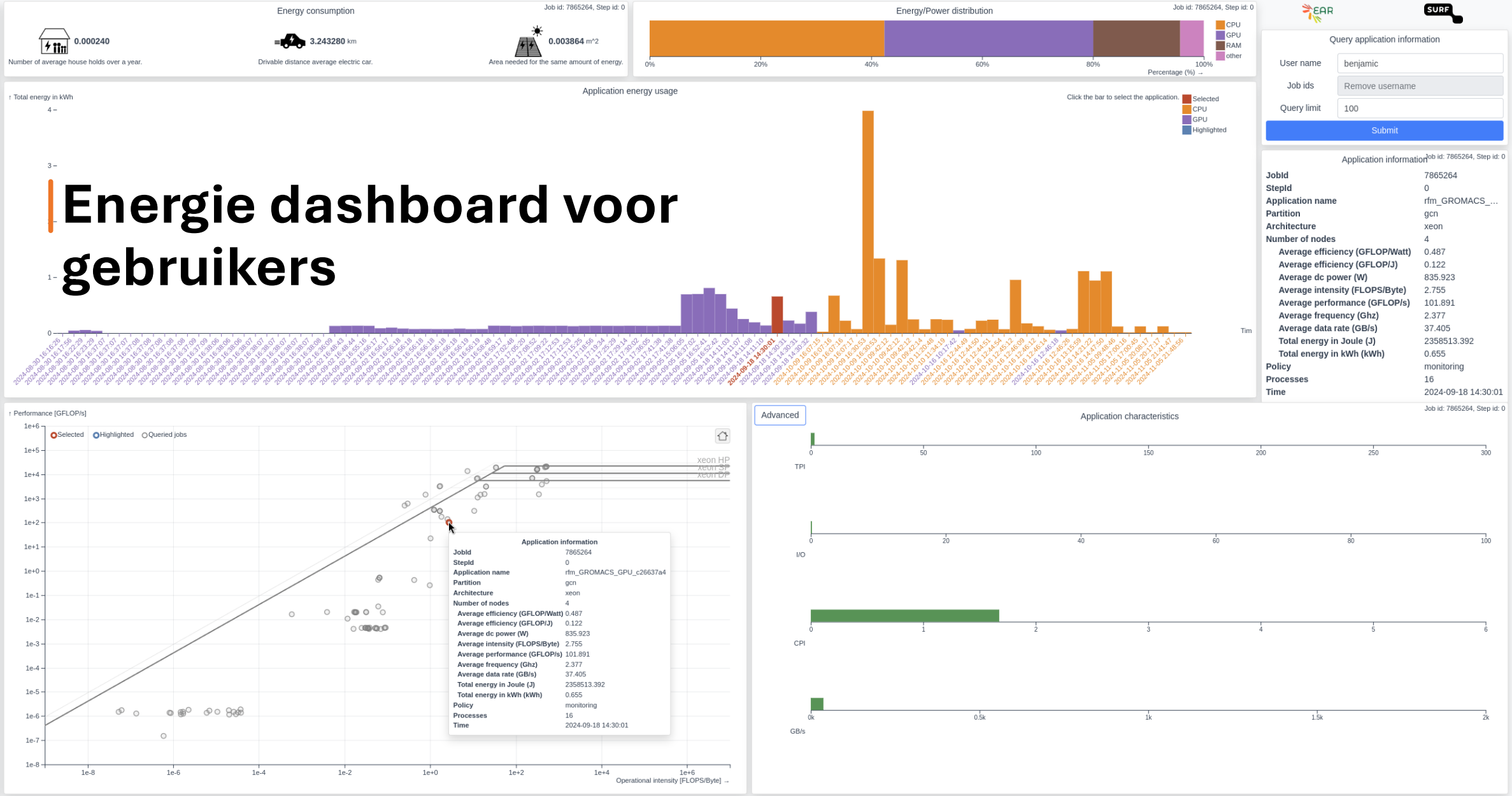
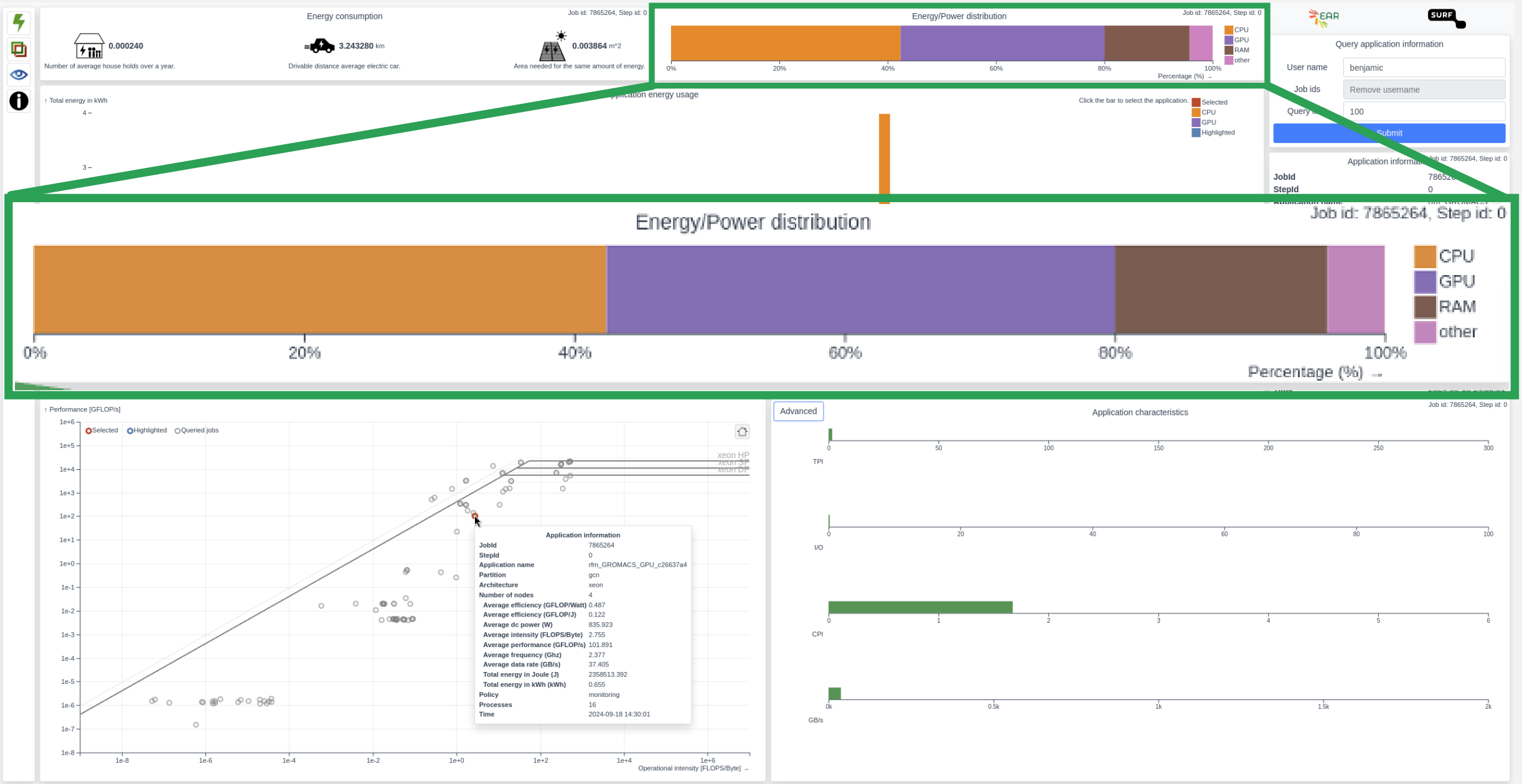


Figure 2 – GPU Energy savings on AI and HPC

https://www.eas4dc.com/

Energie dashboard voor gebruikers



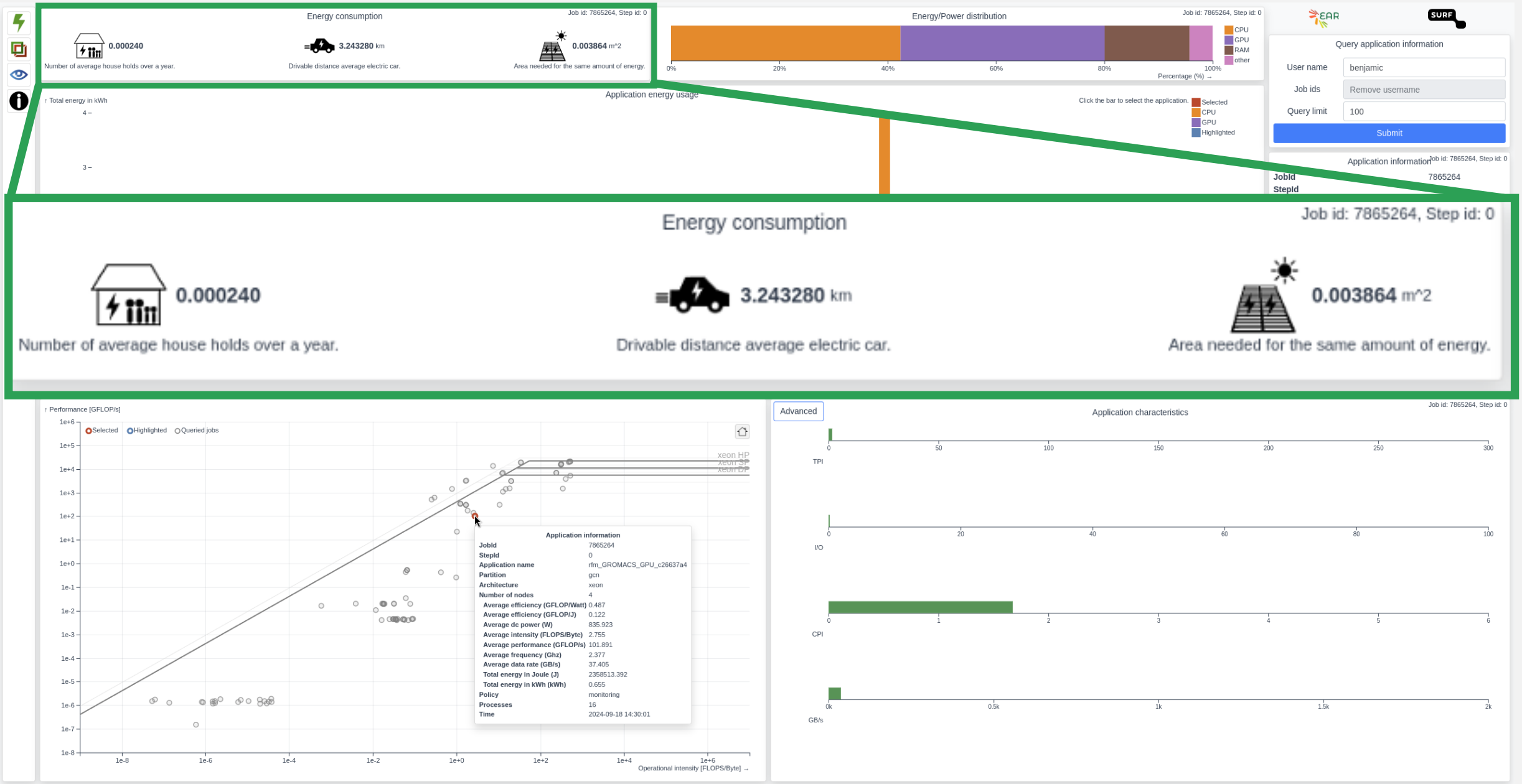

Energy consumption
Job id: 7865264, Step id: 0
0.000240
Number of average house holds over a year.
3.243280 km
Drivable distance average electric car.
0.003864 m^2
Area needed for the same amount of energy.
Energy/Power distribution
Job id: 7865264, Step id: 0
CPU
GPU
RAM
other
0%
20%
40%
60%
80%
100%
Percentage (%) →
EAR
SURF
Query application information
User name
benjamic
Job ids
Remove username
Query
100
Submit
Application information
Job id: 7865264, Step id: 0
JobId
StepId
0
↑ Total energy in kWh
4
3
Application energy usage
Click the bar to select the application.
Selected
CPU
GPU
Highlighted
Energy/Power distribution
Job id: 7865264, Step id: 0
CPU
GPU
RAM
other
0%
20%
40%
60%
80%
100%
Percentage (%) →
↑ Performance [GFLOP/s]
Selected
Highlighted
Queried jobs
xeon HP
xeon SP
xeon DP
Application information
JobId 7865264
StepId 0
Application name rfm_GROMACS_GPU_c26637a4
Partition gcn
Architecture xeon
Number of nodes 4
Average efficiency (GFLOP/Watt) 0.487
Average efficiency (GFLOP/J) 0.122
Average dc power (W) 835.923
Average intensity (FLOPS/Byte) 2.755
Average performance (GFLOP/s) 101.891
Average frequency (Ghz) 2.377
Average data rate (GB/s) 37.405
Total energy in Joule (J) 2358513.392
Total energy in kWh (kWh) 0.655
Policy monitoring
Processes 16
Time 2024-09-18 14:30:01
Operational intensity [FLOPS/Byte] →
Advanced
Application characteristics
Job id: 7865264, Step id: 0
TPI
I/O
CPI
GB/s

Application information

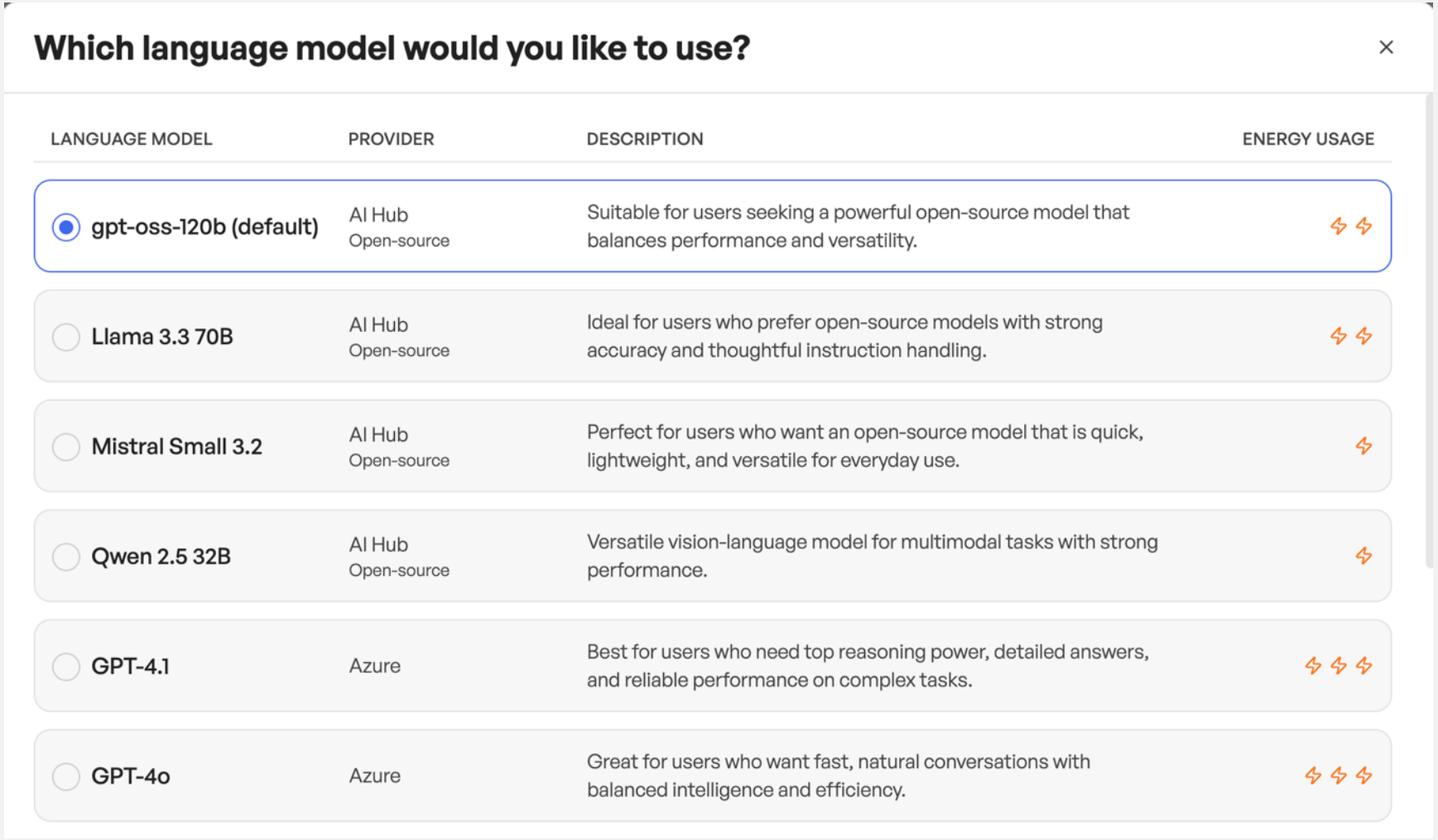
JobId	7865264
StepId	0
Application name	rfm_GROMACS_GPU_c26637a4
Partition	gcn
Architecture	xeon
Number of nodes	4
Average efficiency (GFLOP/Watt)	0.487
Average efficiency (GFLOP/J)	0.122
Average dc power (W)	835.923
Average intensity (FLOPS/Byte)	2.755
Average performance (GFLOP/s)	101.891
Average frequency (Ghz)	2.377
Average data rate (GB/s)	37.405
Total energy in Joule (J)	2358513.392
Total energy in kWh (kWh)	0.655
Policy	monitoring
Processes	16
Time	2024-09-18 14:30:01

1 Gflop/s:

$1x10^9$ Floating Point Operations per second


Energy consumption
Job id: 7865264, Step id: 0
0.000240
Number of average house holds over a year.
3.243280 km
Drivable distance average electric car.
0.003864 m^2
Area needed for the same amount of energy.
Energy/Power distribution
Job id: 7865264, Step id: 0
CPU
GPU
RAM
other
0%
20%
40%
60%
80%
100%
Percentage (%) →
EAR
SURF
Query application information
User name
benjamic
Job ids
Remove username
Query limit
100
Submit
Application energy usage
↑ Total energy in kWh
Click the bar to select the application.
Selected
CPU
GPU
Highlighted
Application information
Job id: 7865264, Step id: 0
JobId 7865264
StepId
Energy consumption
Job id: 7865264, Step id: 0
0.000240
Number of average house holds over a year.
3.243280 km
Drivable distance average electric car.
0.003864 m^2
Area needed for the same amount of energy.
↑ Performance [GFLOP/s]
Selected
Highlighted
Queried jobs
xeon HP
xeon SP
xeon DP
Application information
JobId 7865264
StepId 0
Application name rfm_GROMACS_GPU_c26637a4
Partition gcn
Architecture xeon
Number of nodes 4
Average efficiency (GFLOP/Watt) 0.487
Average efficiency (GFLOP/J) 0.122
Average dc power (W) 835.923
Average intensity (FLOPS/Byte) 2.755
Average performance (GFLOP/s) 101.891
Average frequency (Ghz) 2.377
Average data rate (GB/s) 37.405
Total energy in Joule (J) 2358513.392
Total energy in kWh (kWh) 0.655
Policy monitoring
Processes 16
Time 2024-09-18 14:30:01
Operational intensity [FLOPS/Byte] →
Advanced
Application characteristics
Job id: 7865264, Step id: 0
TPI
I/O
CPI
GB/s

SURF AI-hub: selectie van energie-efficiente LLMs voor specifieke taak

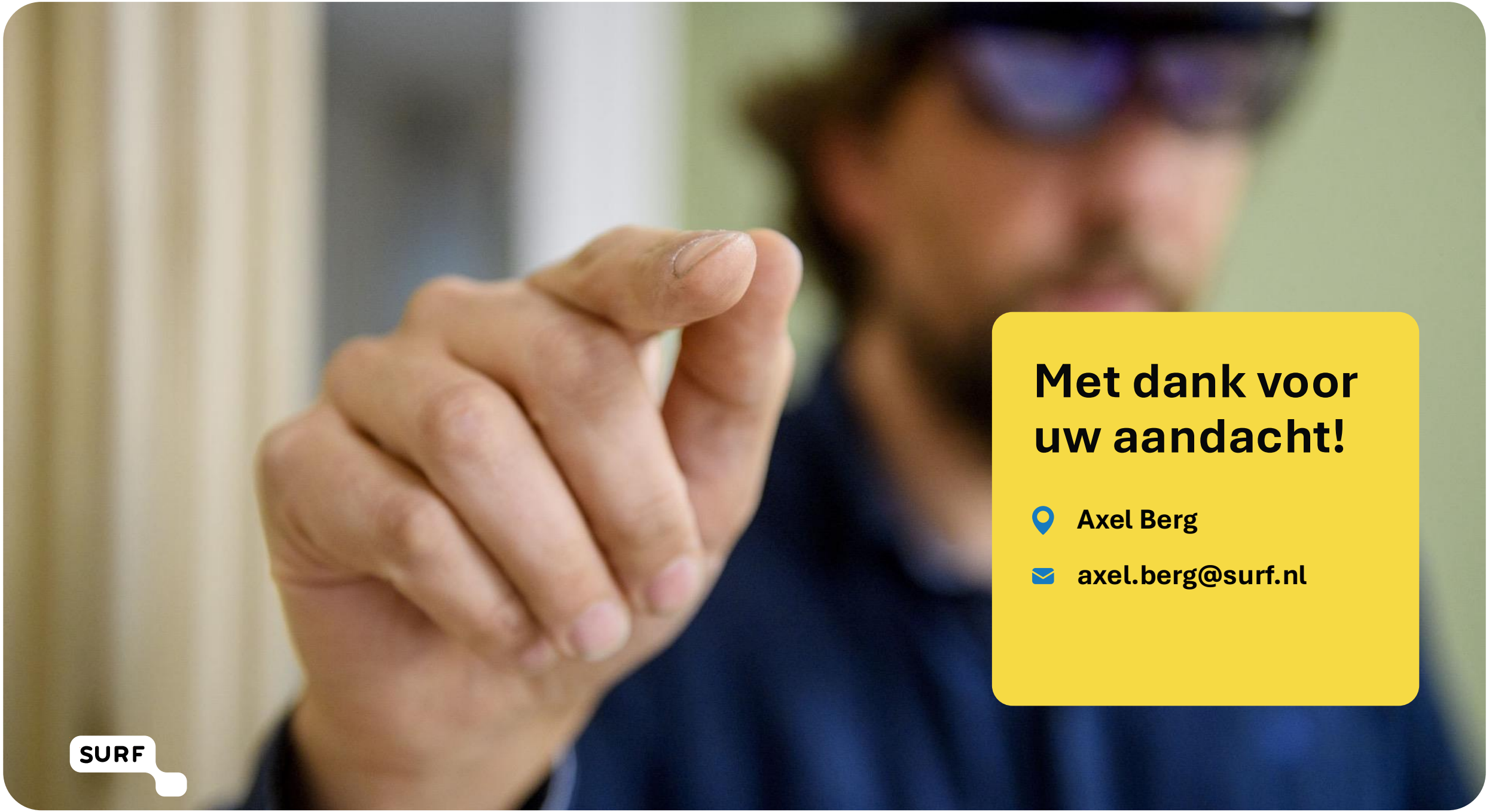


Take away voor vandaag

- **Investeer nu al in de hele keten en leer van elkaar**
  - Energiebronnen
  - Datacenter PUE
  - Hardware
  - Runtime
  - Applicatie
  - Bewustwording

https://www.selfwealth.com.au/blog/how-to-invest-for-exposure-to-renewable-energy

- **Experimenteer met nieuwe technologie**


Met dank voor uw aandacht!
Axel Berg
axel.berg@surf.nl
SURF