EPFL

CONNECTING THE DOTS

ISC High Performance

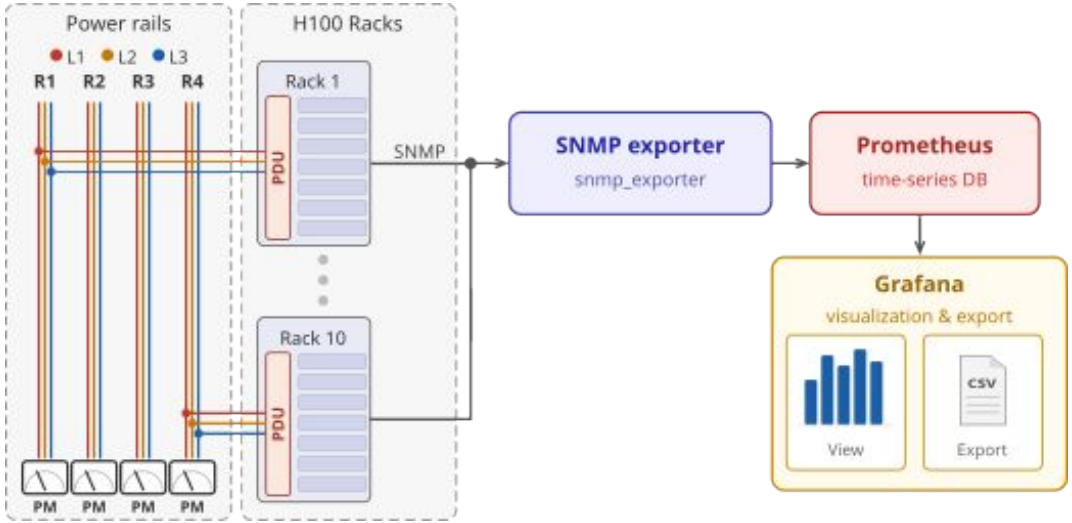
Green500: Adventure in Qualifying a Cluster

Nicolas Richart, Julia Paolini, Xavier Ouvrard, Manuel Cubero-Castan, Gilles Fourestey

Who are we ?

- EPFL, one of the two federal institute of technology in Switzerland
- Work done with three distinct units
  - Operational HPC/AI unit
  - Research unit in sustainable computing
  - Central Sustainability unit
- Gain back knowledge on doing power measurement at Green500 standard
- Validate and enhance power monitoring
- Evaluate environmental impact of IT

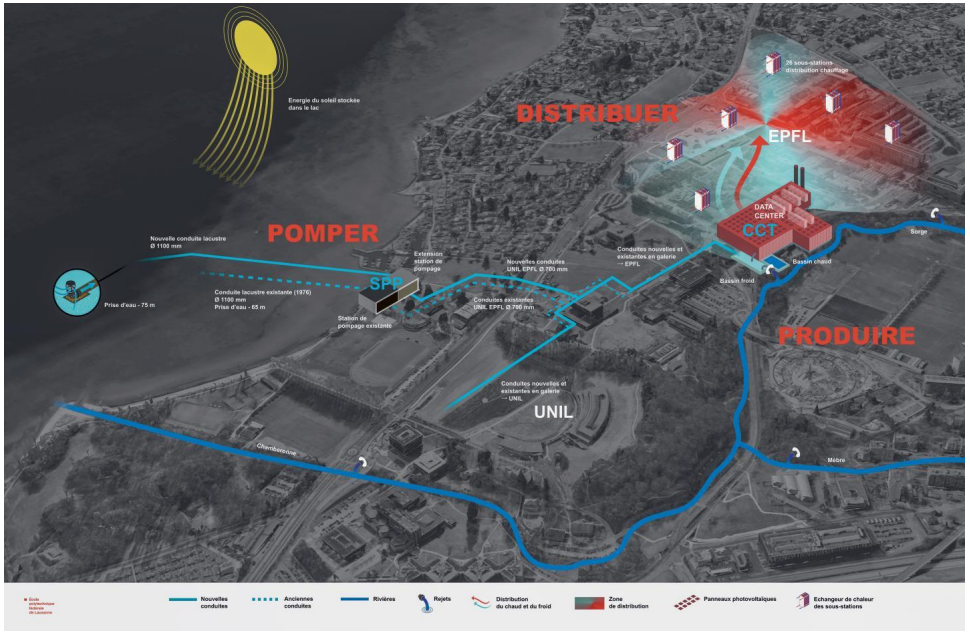
Kuma setup



- 9 racks of H100
  - 84 nodes
  - 80 in TOP500/Green500
- 1 rack of admin
  - network
  - scratch storage
  - front nodes
- 4 power rails
  - 2 direct
  - 2 on UPS
  - Continuously monitored PDUs
- 2 racks of L40s
  - turned off during measurements
- Cooled with water from Lake Geneva
  - Air cooled nodes with water cooled rear doors

Kuma setup

- 9 racks of H100
  - 84 nodes
  - 80 in TOP500/Green500
- 1 rack of admin
  - network
  - scratch storage
  - front nodes
- 4 power rails
  - 2 direct
  - 2 on UPS
  - Continuously monitored PDUs
- 2 racks of L40s
  - turned off during measurements
- Cooled with water from Lake Geneva
  - Air cooled nodes with water cooled rear doors

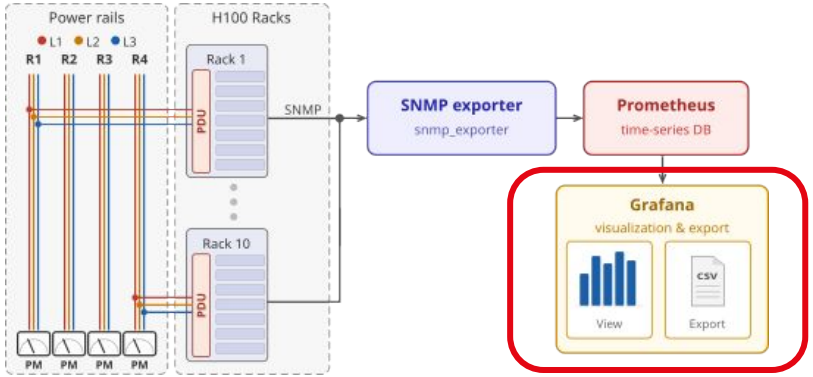


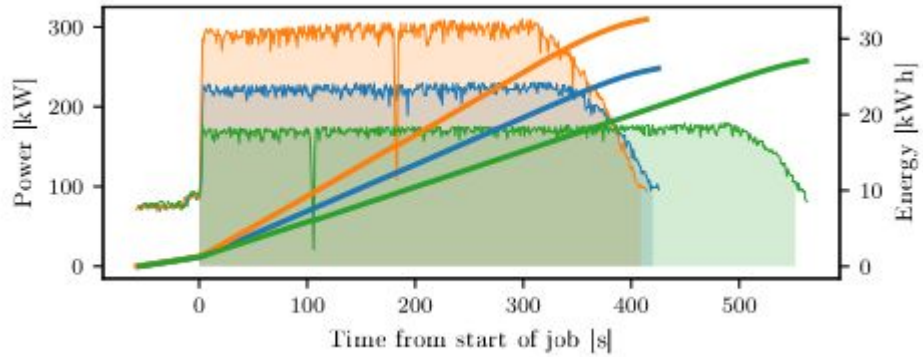
What has been done on the Kuma cluster

- Submission TOP500 and Green500 November 2024
  - Level 2 - 55 GFlop/s/Watt
  - 102nd on TOP500 and 23rd on Green500
- More in depth tests of the cluster to better characterize the machine
- Submission TOP500 and Green500 November 2025
  - Level 3 - 53 GFlop/s/Watt
  - Applying MITSI (Methodology for IT Services environmental Impact assessment)
    - CO_2 impact
    - Only HPL runs

Level 2 measurements

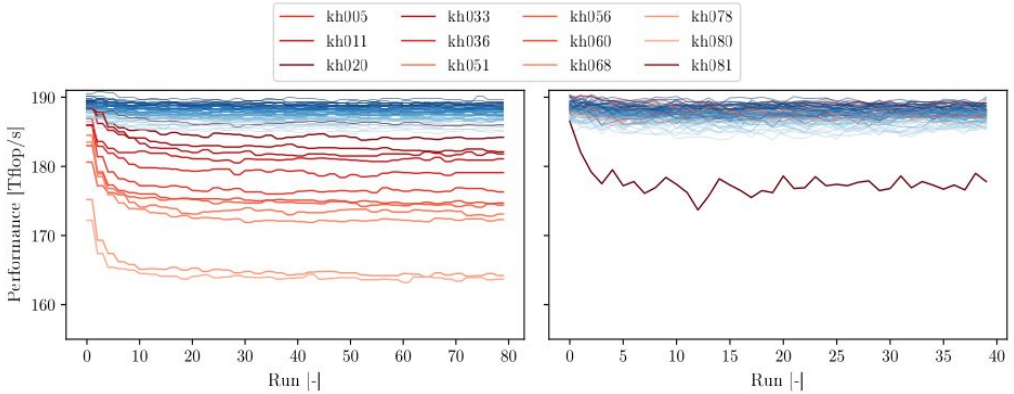
- Power from the PDUs
  - only 4kHz sampling
- Exported from Grafana
- Cold runs
  - let the nodes cool down between runs
- Power capping of the GPUs
  - 350W, 500W, 700W
  - Highest power efficiency at 500W





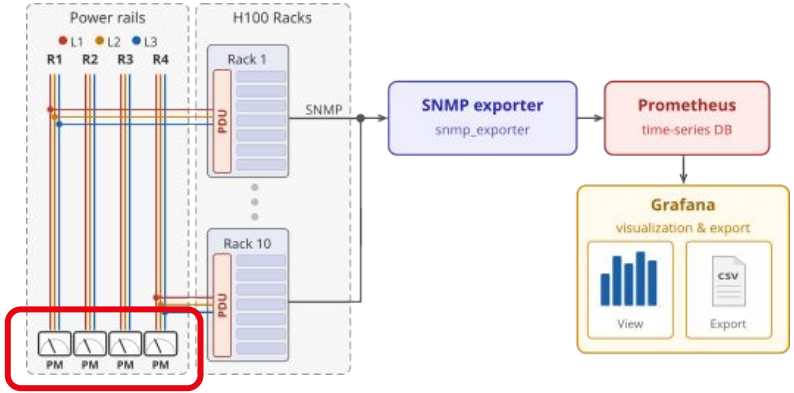
Investigations before level 3

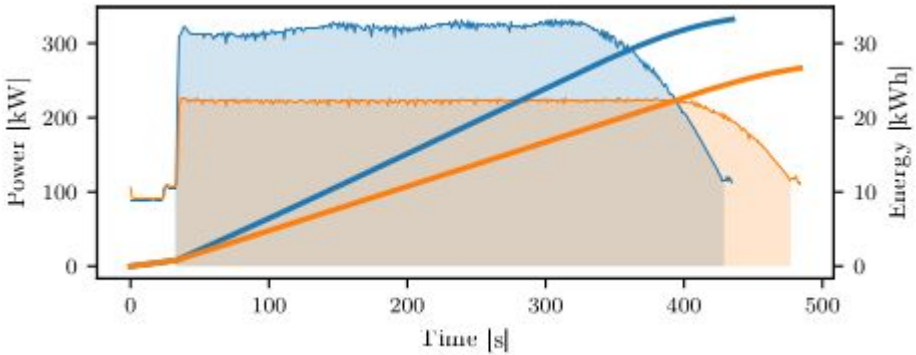
- The questions we had:
  - Should we let the machine cool down between the runs
  - Is there a network imbalance
  - Is the rack cooling sufficient
- One node loop test
  - Drop of performance after 10 runs on some nodes
  - Thermal paste issue
- All pair of nodes tested
  - No imbalance measured
- A full rack test with extra temperature sensors
  - Stayed in the expected range



Level 3 measurements

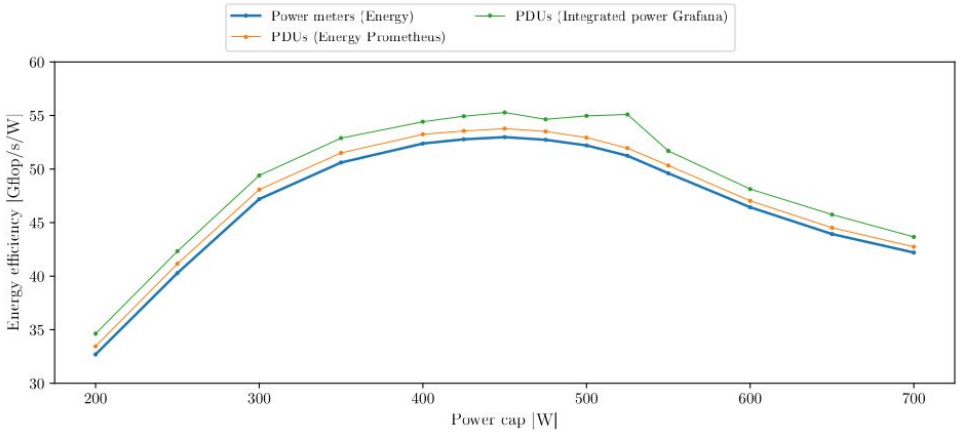
- Energy from Power meters on each rails
  - more than 5kHz
- Hot runs
  - no need for cool down time
- Power capping of the gpus
  - every 50W steps from 250W to 700W
  - Highest power efficiency at 450W cap





Level 2 vs level 3 setups

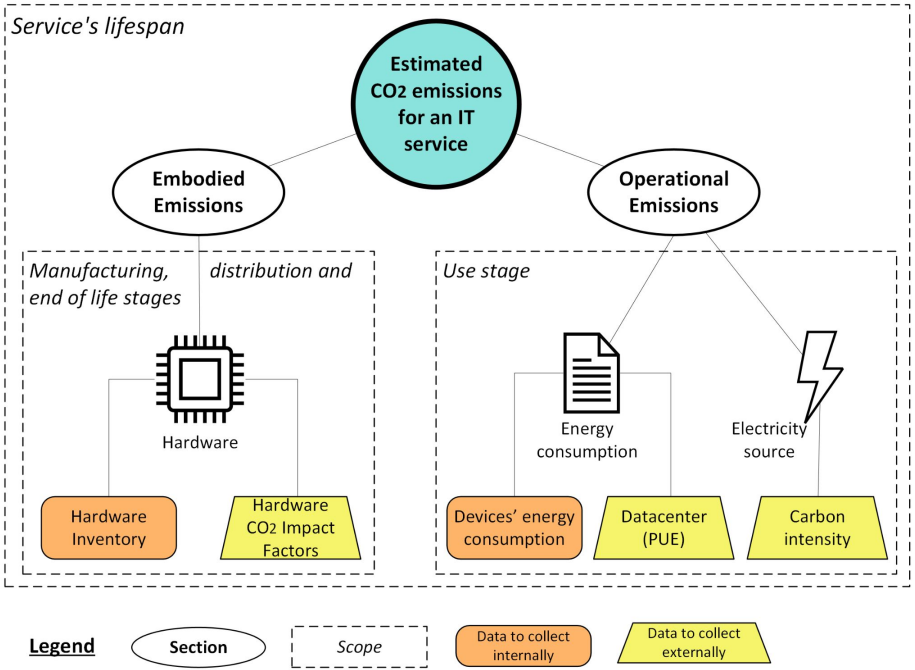
- Local monitoring limitations
  - PDUs data harder to process
  - Timing issue
    - Ingestion date
- Power meters measurements
  - Need electricians to install/remove
  - Time offset issue:
    - 2s between direct and UPS rails



	GFlop/s/W	%
PDUs (Grafana)	55.2	-
PDUs (Prometheus)	53.7	-3%
Power meters	53.0	-4%

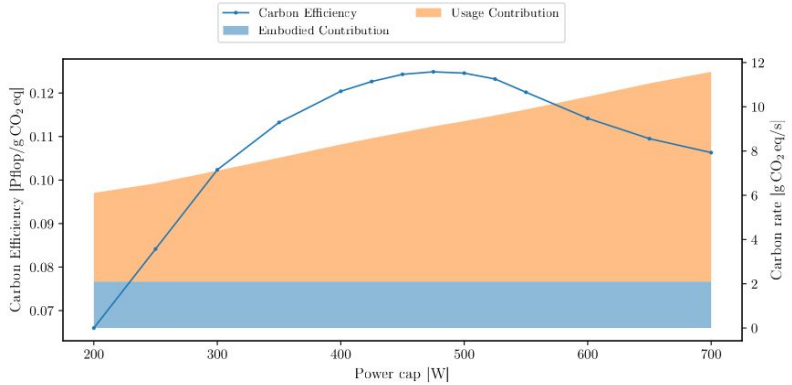
MITSI

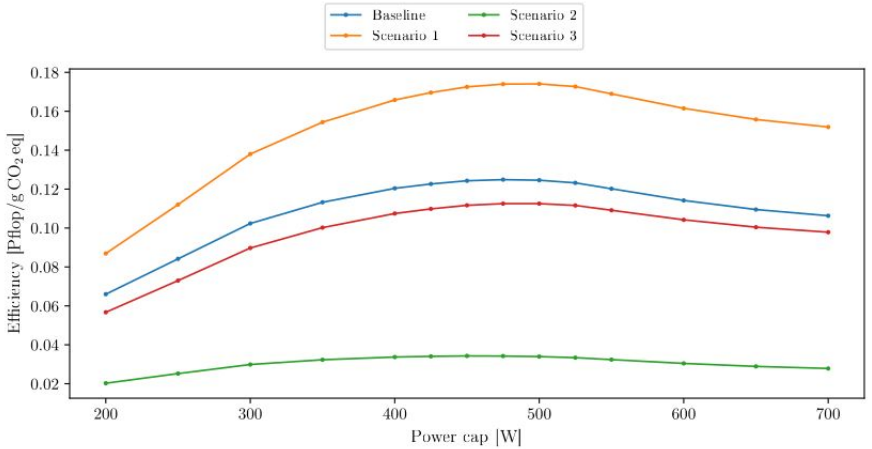
- Embodied emissions
  - What is accounted
    - CPUs
    - GPUs
    - RAM
    - PSU
    - Storage
    - Network
  - 387 tCO_2eq
  - 23 $gCO_2eq/hGPU$ (6 years)
- Usage emissions
  - Average power usage over a representative time period
    - Real core time for HPL
  - PUE
    - 1.3
  - Carbon Intensity (CI)
    - 90 gCO_2eq/kWh (2025)



Carbon efficiency

- $$C_{\text{eff}} = \frac{Perf}{C_{\text{embodied}} + C_{\text{usage}}} = \frac{Perf}{C_{\text{embodied}} + (\bar{P} \cdot CI \cdot PUE)}$$
- Highest carbon efficiency at a capping of 475W
- Variation on the measured values
  - Scenario 1:
    - CI from 2024 (57 gCO_2eq/kWh)
  - Scenario 2:
    - Decreasing life time to 4 years
  - Scenario 3:
    - Using a higher CI (400 gCO_2eq/kWh)





Conclusion

- Continuous power monitoring following EE HPC WG recommendations in place
- Carbon impact highly sensitive to location
- Future works:
  - Need to run more types of benchmarks (not only HPL)
  - Need to consider more than just carbon
  - Add carbon efficiency measurements to Green500
    - As informative values more than a ranking
    - Carbon500 ?