CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

Adastra energy

B. Cirou, **Journée « Energie, infrastructure et calcul, quel équilibre ? »**

Montpellier - 27 mars 2025

CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

Layout

- ❑ Adastra Architecture
- ❑ Metrics
- ❑ Data
- ❑ Work in progress

ADASTRA Architecture

HPE Cray EX4000

- **544 Genoa nodes**

2 x 9654 @ 2.4GHz

- **356 MI250X nodes**

8 x GFX90A @ 1.5GHz

3rd Green 500 (2023)

- **28 MI300A nodes**

4 x GFX942 @ 1.8GHz

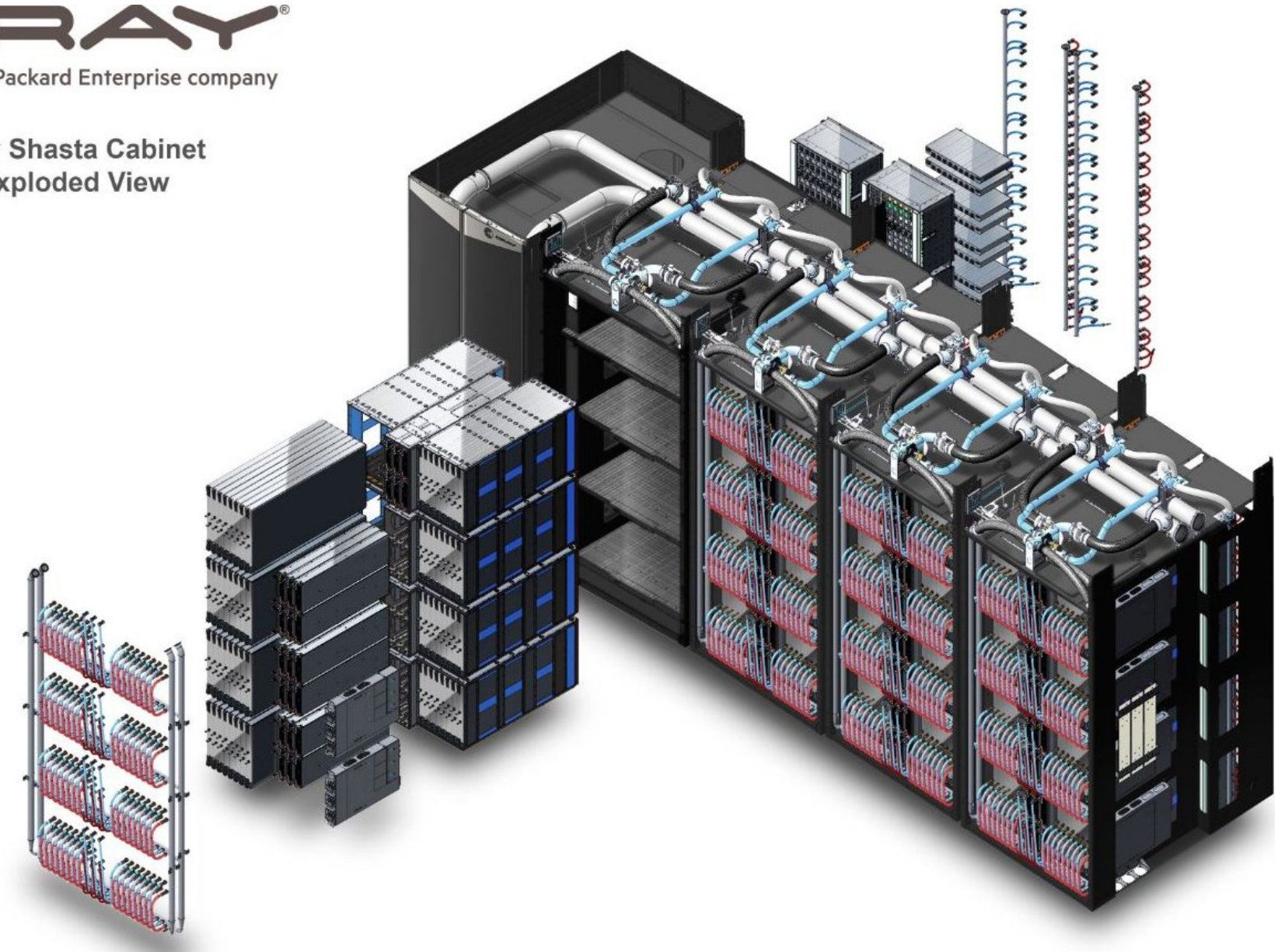
3rd Green 500 (2024)

ADASTRA Architecture

CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

CRAY®
a Hewlett Packard Enterprise company

Cray Shasta Cabinet
Exploded View

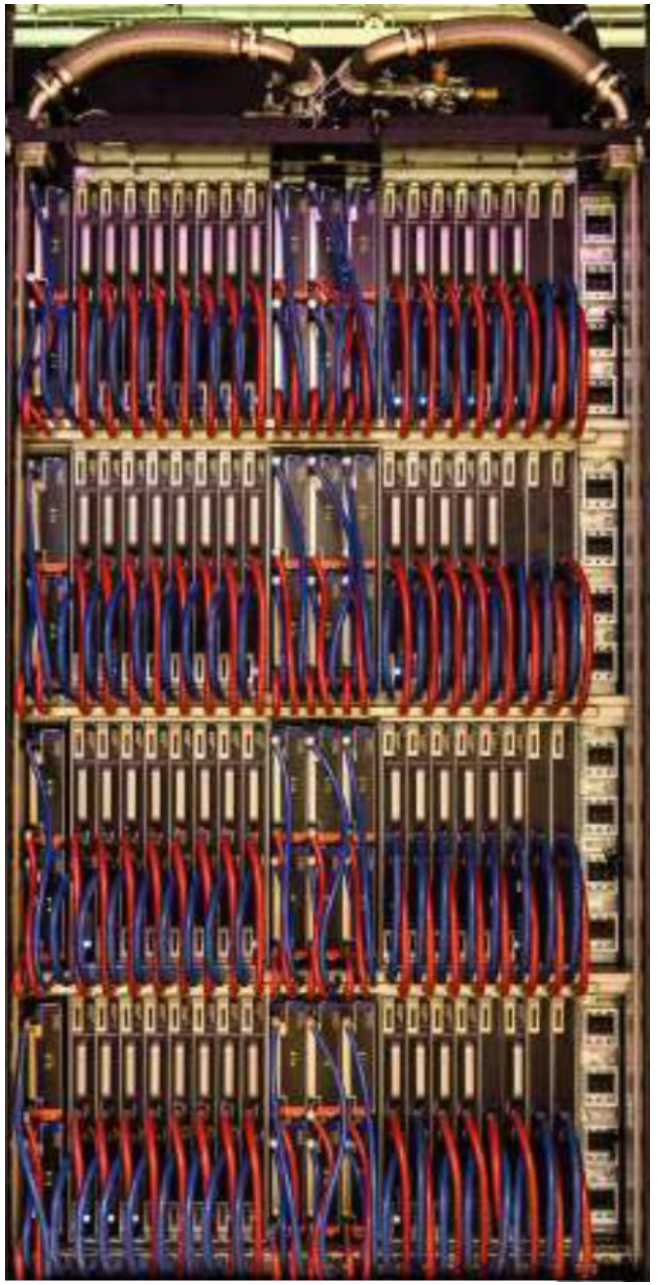


ADASTRA Architecture

Cooling

- **97% watercooled**
- **1493 kW peak** (MI250X + GENOA blades)

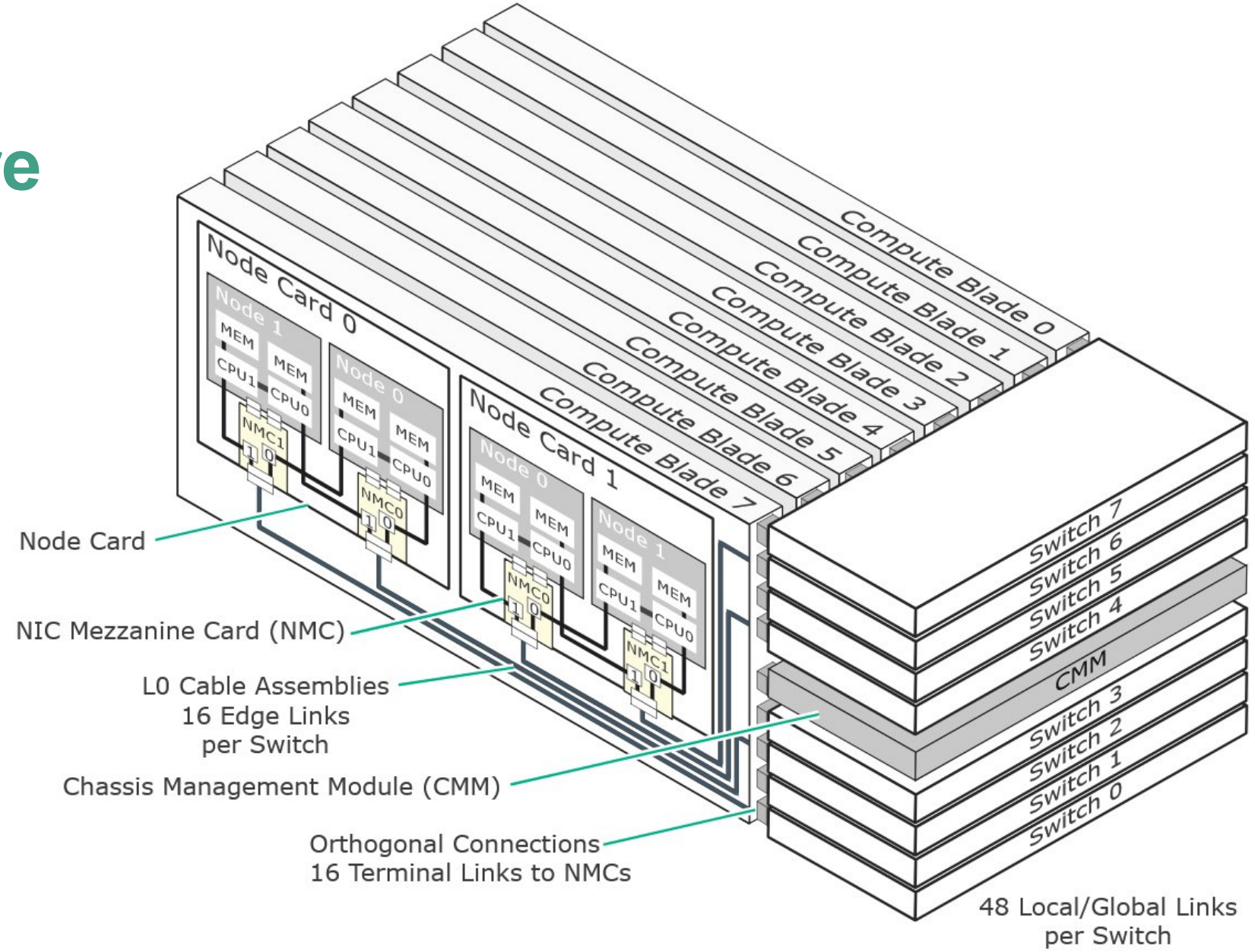
- CDU **Inlet** temp: **31°C**
- CDU **Outlet** temp: **42°C**
- Flow: **90 m3/h**
- Pressure drop: **160 kPa**



ADASTRA Architecture

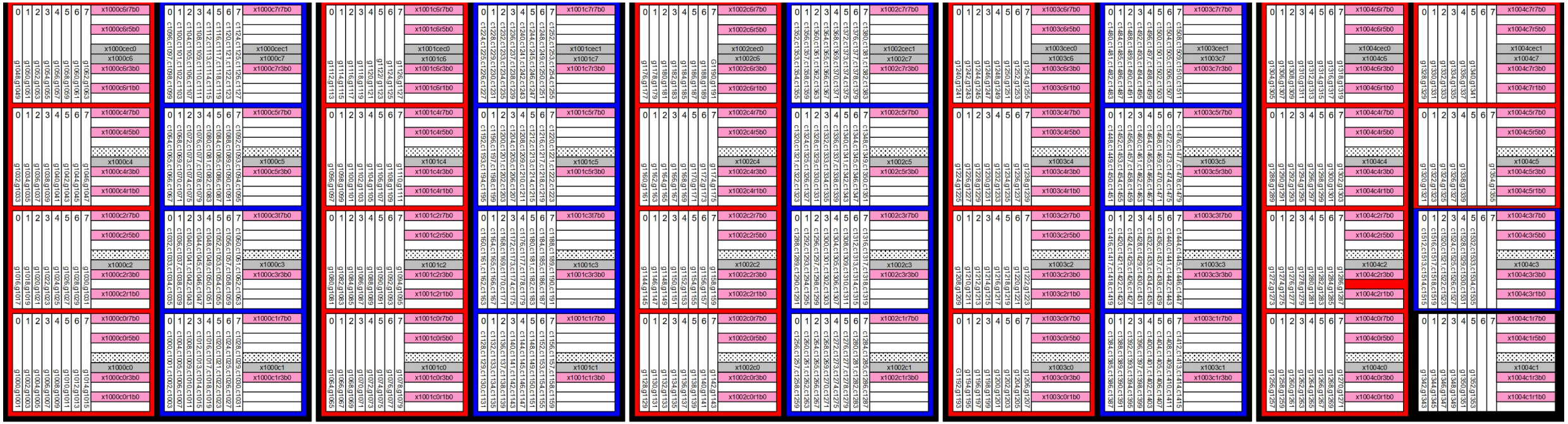
Genoa enclosure

Direct Liquid Cooling



ADASTRA Architecture

Mixed nodes MI250X / Genoa cluster



Nœuds GPU Nœuds CPU Commutateurs Slingshot CEC & CMM

METRICS

Monitoring: Cray HPCM (Kibana)

CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

METRICS

Admin tools : Cray HPCM

Get Node Energy

ncn-w001# cray capmc get_node_energy create --nids *NID_LIST* \ --start-time '2020-03-04 12:00:00' --end-time '2020-03-04 12:10:00' --format json

Get Node Energy Stats

ncn-w001# cray capmc get_node_energy_stats create --nids *NID_LIST* \ --start-time '2020-03-04 12:00:00' --end-time '2020-03-04 12:10:00' --format json

Get Node Energy Counter

ncn-w001# cray capmc get_node_energy_counter create --nids *NID_LIST* \ --time '2020-03-04 12:00:00' --format json

Get Node Power Control and Limit Settings

ncn-w001# cray capmc get_power_cap create –-nids *NID_LIST* \ --format json

Get System Power

ncn-w001# cray capmc get_system_power create \ --start-time '2020-03-04 12:00:00' --window-len 30 --format json
ncn-w001# cray capmc get_system_power_details create \ --start-time '2020-03-04 12:00:00' --window-len 30 --format json

Get Power Capping Capabilities

ncn-w001# cray capmc get_power_cap_capabilities create –-nids *NID_LIST* \ --format json

Set Node Power Limit

ncn-w001# cray capmc set_power_cap create –-nids *NID_LIST* \ --node 225 --format json

Remove Node Power Limit (Set to Default)

ncn-w001# cray capmc set_power_cap create –-nids *NID_LIST* \ --node 0 --format json

CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

METRICS

Node metrics: Cray PM Counters

/sys/cray/pm_counters

- **power**: Point-in-time power (Watts).
- **energy**: Accumulated energy, in joules.
- **cpu_power**: Point-in-time power (Watts) used by the CPU domain.
- **cpu_energy**: The total energy (Joules) used by the CPU domain.
- **cpu_temp**: Temperature reading (Celsius) of the CPU domain.
- **memory_power**: Point-in-time power (Watts) used by the memory domain.
- **memory_energy**: The total energy (Joules) used by the memory domain.
- **accel_energy**: Accumulated accelerator energy (Joules).
- **accel_power**: Accelerator point-in-time power (Watts).
- **raw_scan_hz**: The power management scanning rate for all data in pm_counters.
  […]

METRICS

Job scheduler: SLURM

SLURM configuration

AcctGatherEnergyType=acct_gather_energy/pm_counters
*JobAcctGatherFrequency=task=30,**energy**=30,filesystem=30*

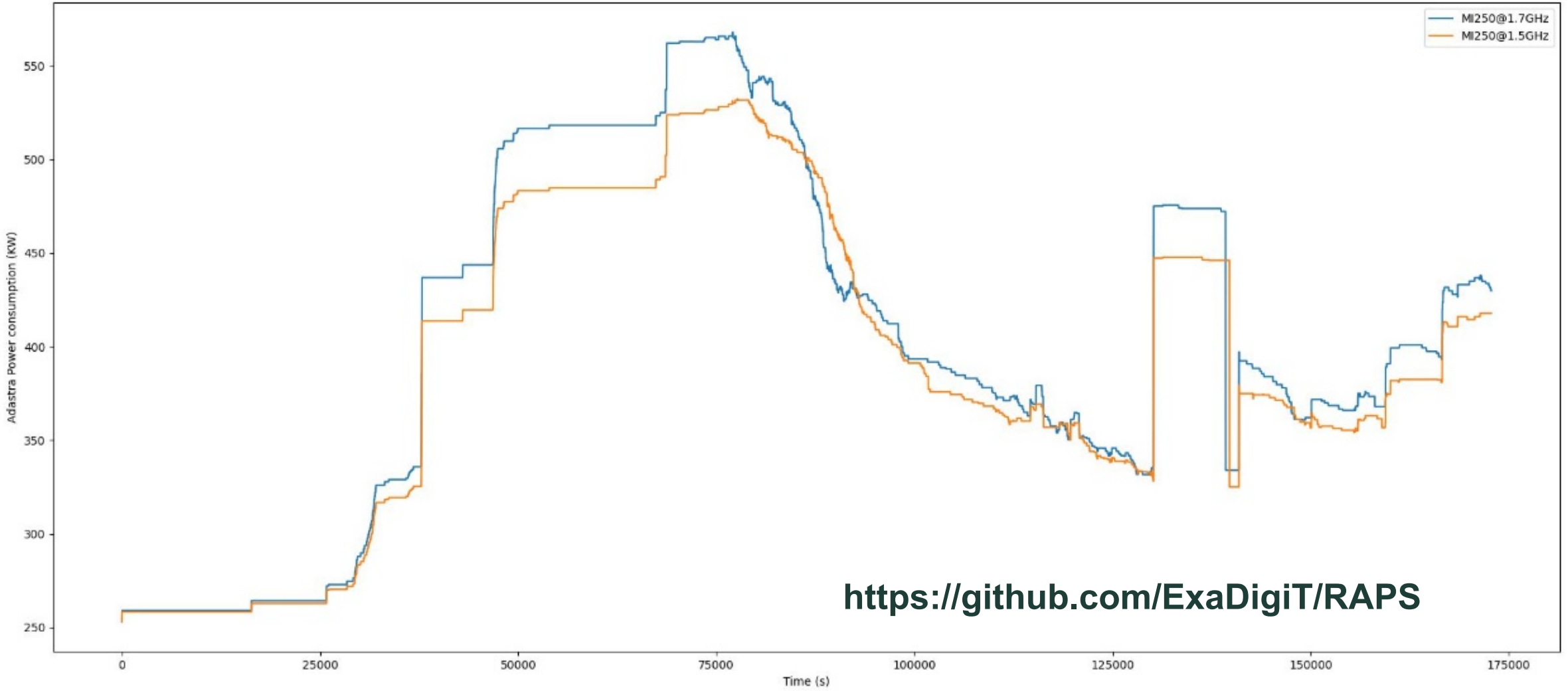
Frequency and power capping

- ❑ Defaults on ADASTRA MI250 nodes
  - **Turbo ON on Trento CPU**
  - **Frequency capping to 1500 MHz on MI250x GPU**
- ❑ Special SLURM plugin for dedicated job tuning (SPANK)
  - **Power capping management**
    MI250X : 100 - 560W
    MI300A : 100 – 550W
    #SBATCH --gpu-power-cap=430
  - **Frequency capping management**
    MI250X : 500 - 1700 MHz
    MI300A : 500 – 1800 MHz
    #SBATCH --gpu-srange=500-1400

METRICS

PM_counter based (RAPS simulation)

Theoretical pure MI250X racks



https://github.com/ExaDigiT/RAPS

METRICS

Job scheduler: SLURM

- Consumption of current running job :

sstat -j <jobid> --format=JobID,ConsumedEnergy –P JobID|ConsumedEnergy

<jobid>.<current step>|3.20K

- Consumption of ended jobs :

sacct -j <jobid> --format=JobID,ConsumedEnergy –P JobID|ConsumedEnergy

<jobid>.<current step>|3.20K

- Job report :

CINES Job Report:

----

o Estimated energy consumption: ***254076 Joules***

(representing ***~ 27% of the maximum nodes utilization****)*

CINES
CENTRE INFORMATIQUE NATIONAL DE L'ENSEIGNEMENT SUPÉRIEUR

METRICS

CINES Data Cluster (ELK, VictoriaMetrics, …)

- Data Lake
- Data Warehouse
- Data Mart

- **Jobs Data Warehouse**
  - Times
  - Node list
  - Job Energy
  - GPU energy
  - Maxrss
  - Modules loaded
  - Pm counters
  - …

Data

Published October 29, 2024 | Version 1.0

Adastra jobs MI250 15days

Cirou (Producer)[1]

Adastra jobs MI250 15days

Files

Files (1.2 MB)

Name

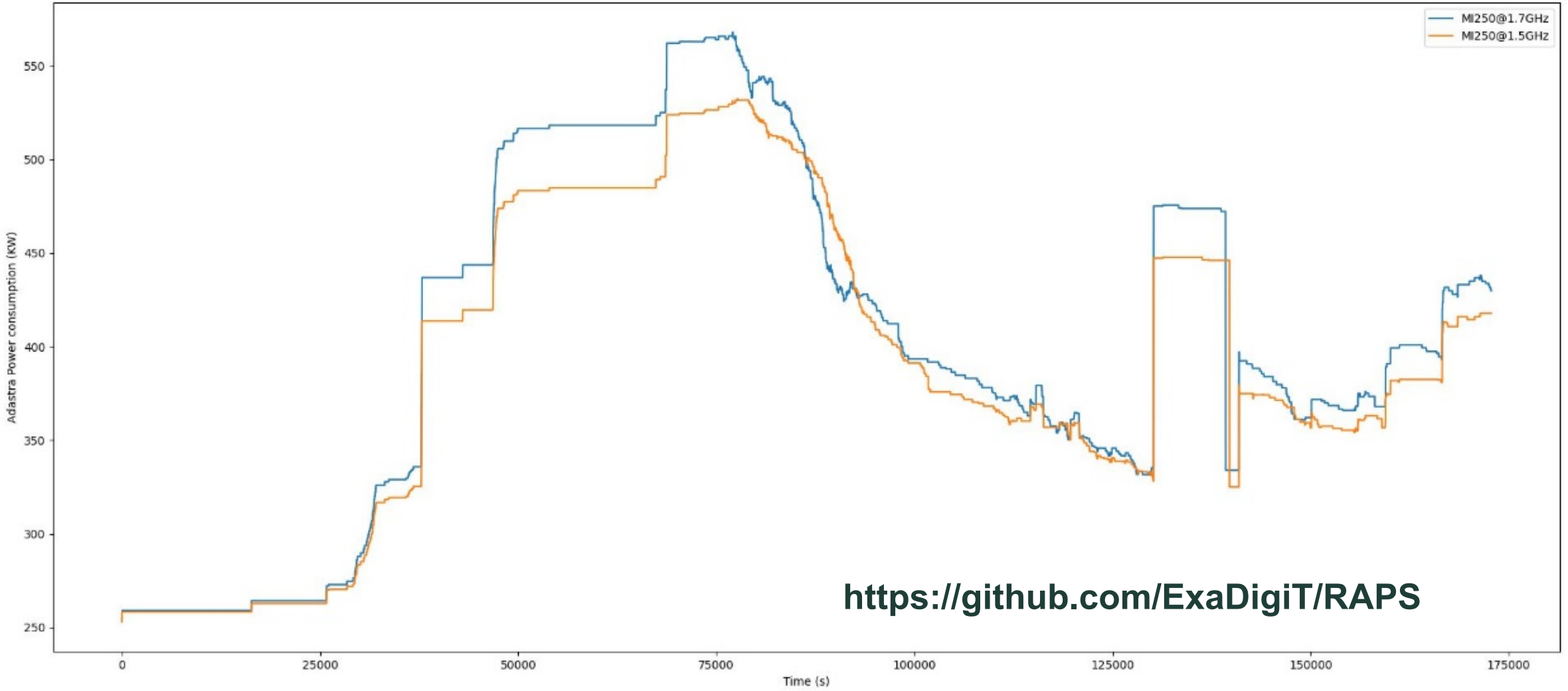
AdastaJobsMI250_15days.parquet

md5:0e875eacc1c9f31426715744b3da754e

Data

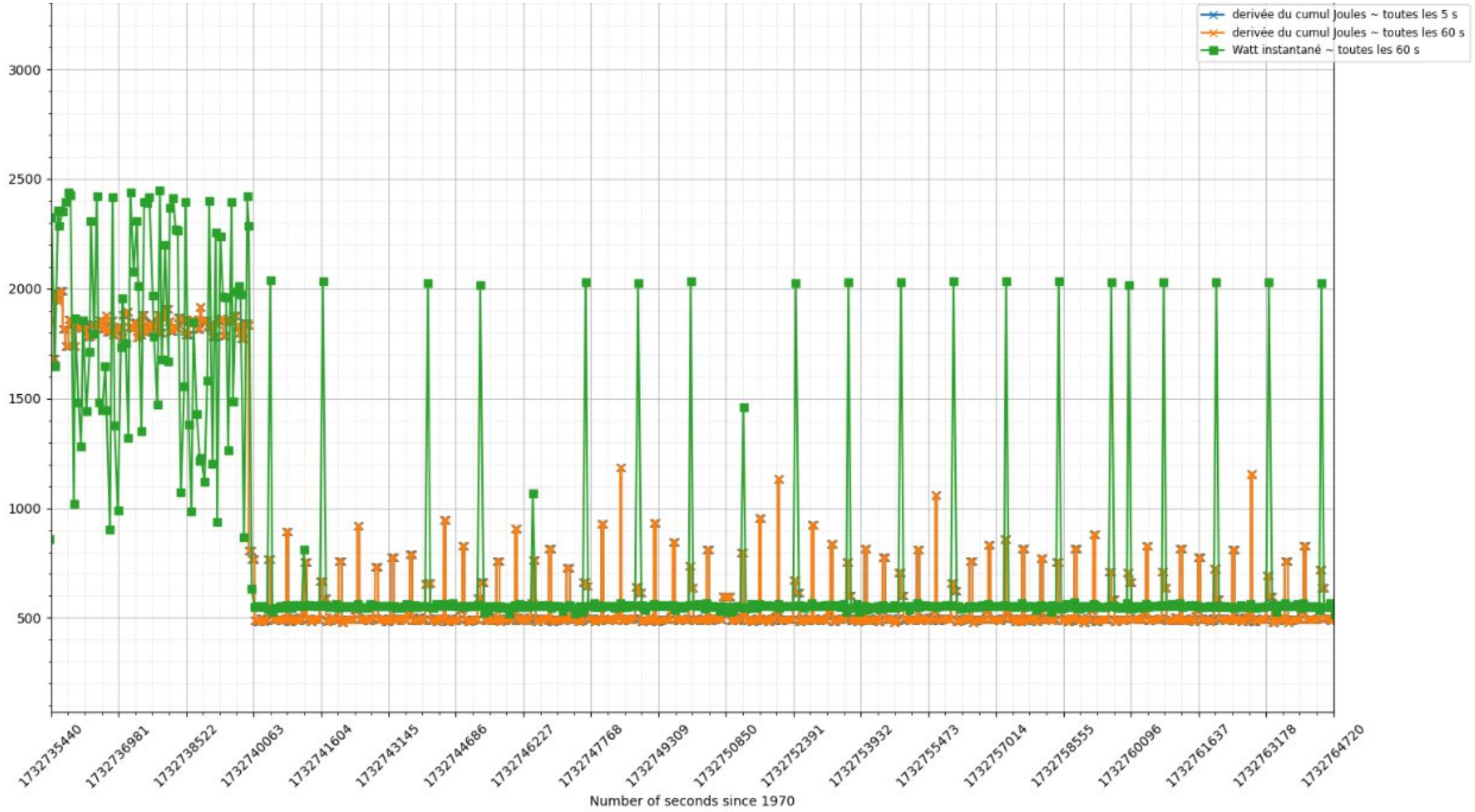
PM_counter based (RAPS simulation)

Theoretical pure MI250X racks



https://github.com/ExaDigiT/RAPS

Data


derivée du cumul Joules ~ toutes les 5 s
derivée du cumul Joules ~ toutes les 60 s
Watt instantané ~ toutes les 60 s
3000
2500
2000
1500
1000
500
1732735440
1732736981
1732738522
1732740063
1732741604
1732743145
1732744686
1732746227
1732747768
1732749309
1732750850
1732752391
1732753932
1732755473
1732757014
1732758555
1732760096
1732761637
1732763178
1732764720
Number of seconds since 1970

Work in progress

Digital Twin project:

EXADIGIT

Participating Organizations

OAK RIDGE National Laboratory
CSC
epcc
JÜLICH Forschungszentrum
KTH VETENSKAP OCH KONST
VSB TECHNICAL UNIVERSITY OF OSTRAVA | IT4INNOVATIONS NATIONAL SUPERCOMPUTING CENTER
CINES
pawsey
ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA
TEXAS STATE UNIVERSITY

Industry Partners

Hewlett Packard Enterprise
NVIDIA

Work in progress

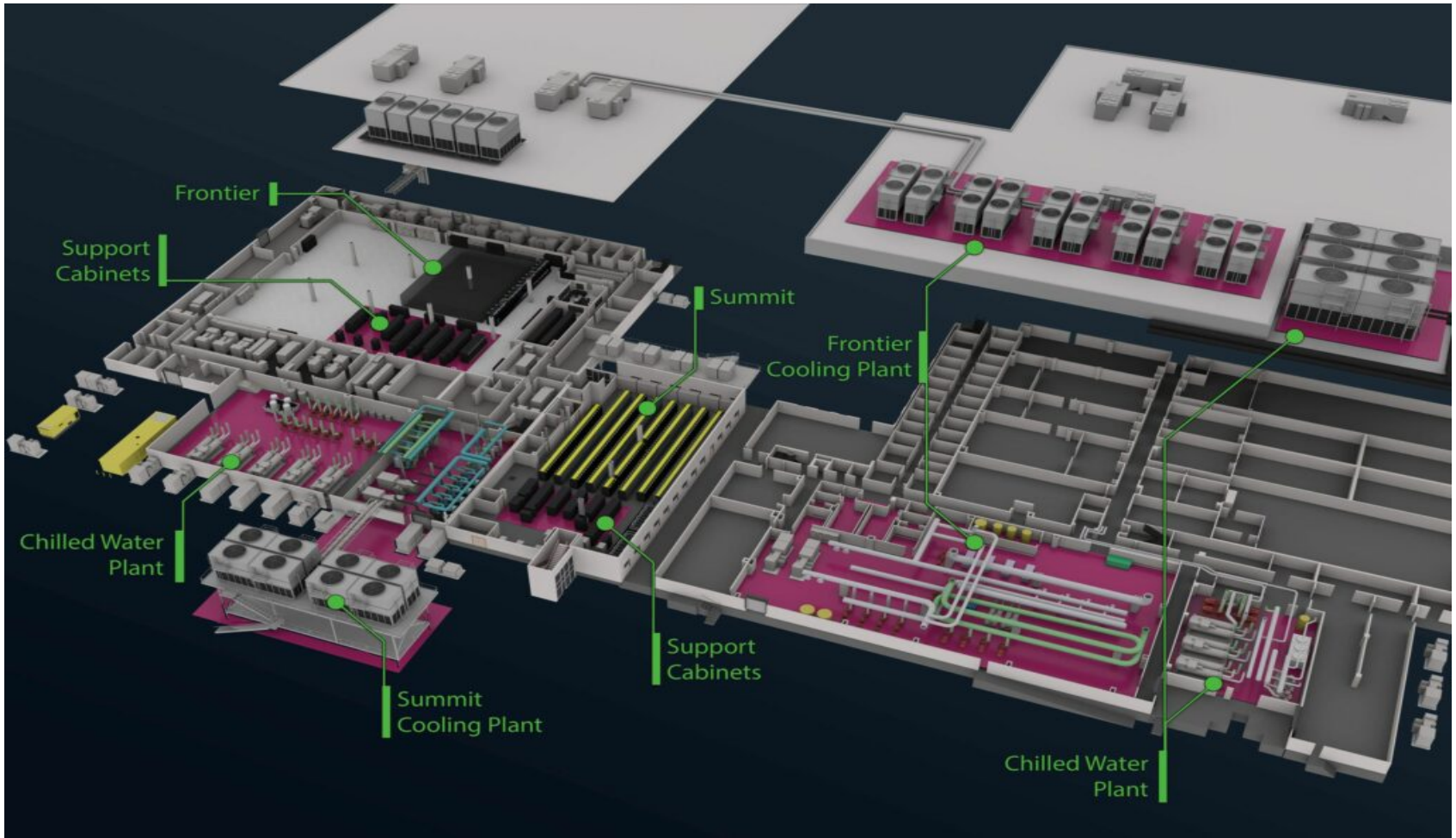
RAPS: Resource Allocator and Power Simulator

RAPS module can replay workloads from telemetry, reschedule them, or simulate synthetic workloads on the supercomputer to analyze the resulting energy consumption; further details are provided in Section

Work in progress

Modelica-based thermo-fluids cooling mode

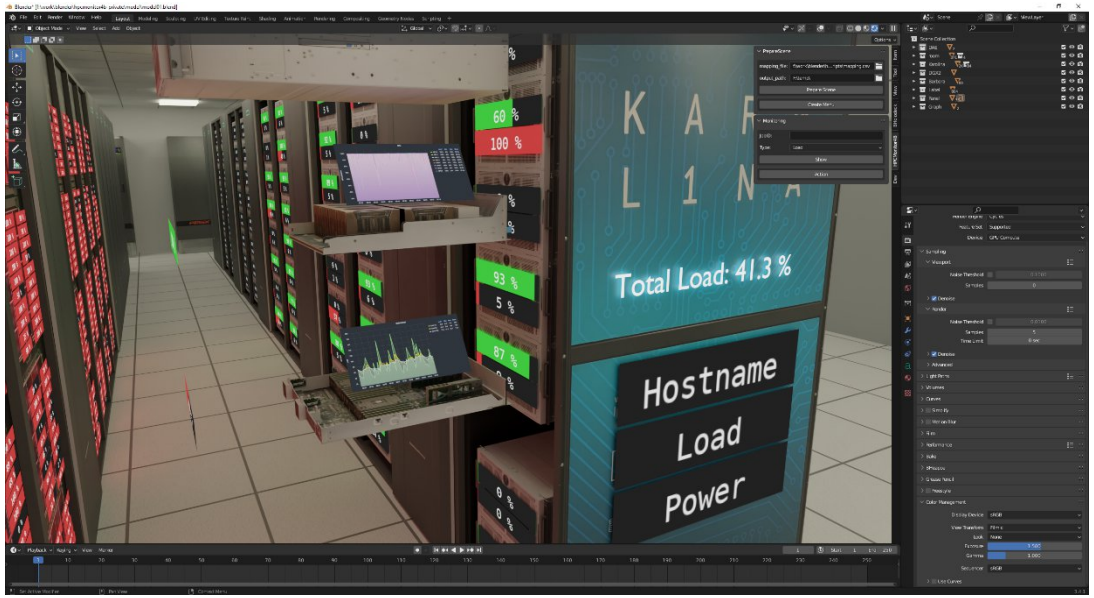
OpenModelica / Dymola®



Work in progress

Visualization

HPCMonitor4B
https://code.it4i.cz/blender/hpcmonitor4b



UnrealEngine

https://www.youtube.com/watch?v=SLWVqEonVgw