Enhancing Job Monitoring with Power Consumption Metrics

Natalia Szczepanek, Domenico Giordano
(CERN)

WLCG Workshop
08/05/2025

Agenda

1. Intro - Objective
2. Infrastructure and methodology
3. Power Measurements in pilot sites
   - DESY
   - AGLT2
   - Glasgow
4. Next Steps

Natalia Szczepanek (CERN)
Domenico Giordano (CERN)

Objective

Extend WLCG job monitoring to include node-level power consumption data, enabling energy consumption analysis at job level

Demonstrate feasibility through a prototype using HEP Benchmark Suite, paving the way for community-wide R&D and adoption.

Work performed in the context of the HEPiX Benchmarking Working Group

Special thanks to the teams at the three sites for their prompt and supporting efforts:

AGLT2: Shawn McKee, Wendy Dronen
DESY: Jan Hartmann, Thomas Hartmann
Glasgow: Emanuele Simili

Reminder

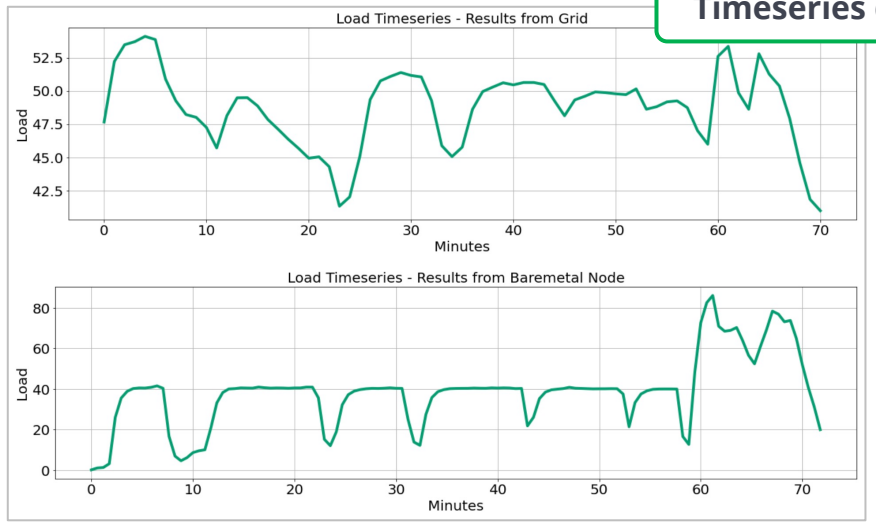
HEP Benchmark Suite

- The Suite is an orchestrator of multiple benchmarks. The suite incorporates metrics such as machine load, memory usage, memory swap, and notably, **power consumption***
- Suite Plugins:
  - Run alongside benchmarks
  - Flexible modification and addition of collected metrics
- Only requirement: The command needs to be executed by the suite user

Configuration

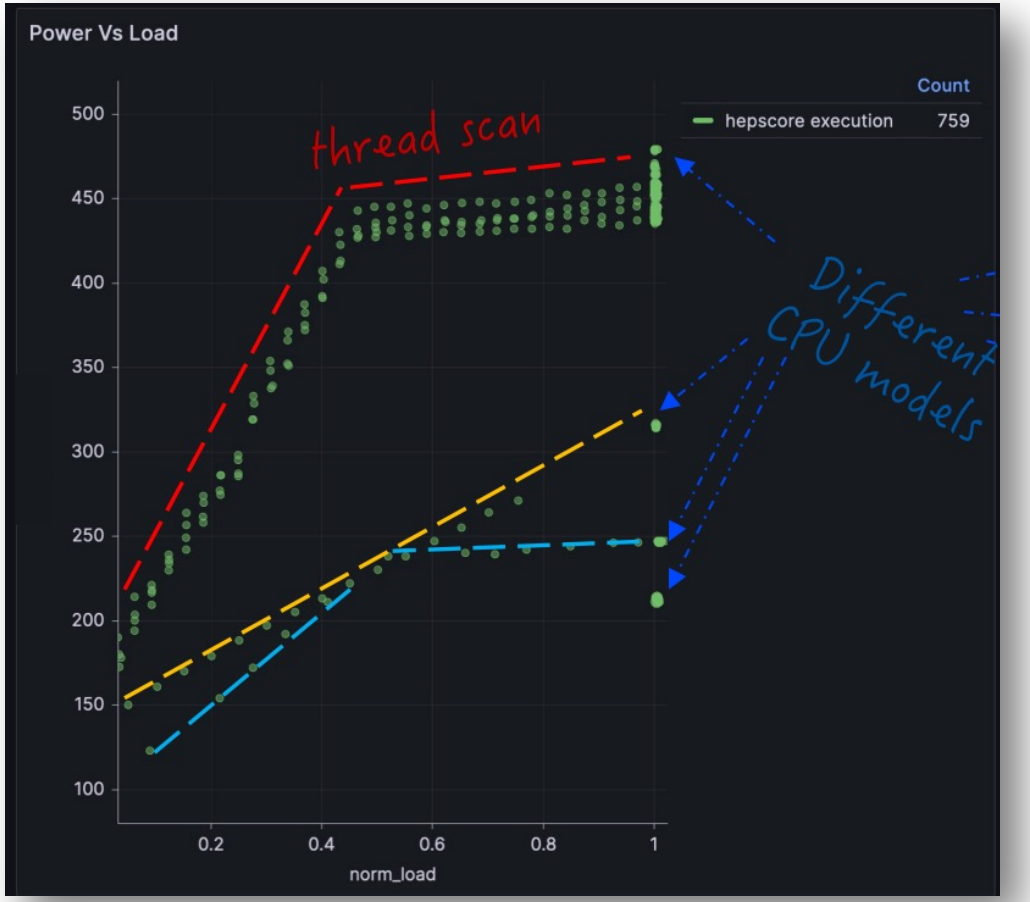
```
plugins:
  CommandExecutor:
    metrics:
      cpu-frequency:
        command: >-
          cpupower -c all frequency-info -f | grep 'current CPU frequency:' |
          grep -o '[0-9]\{7,\}' | awk '{s+=$1; c++} END {print (s/c)/1000}'
        interval_mins: 1
        regex: (?P<value>\d+.\d+).*
        unit: MHz
      load:
        command: uptime
        interval_mins: 1
        regex: 'load average: (?P<value>\d+.\d+),'
        unit: ''
      power-consumption:
        command: ipmitool dcmi power reading
        description: >-
          Retrieves power consumption of the system. Requires elevated
          privileges.
        interval_mins: 1
        regex: 'Instantaneous power reading:\s*(?P<value>\d+) Watts'
        unit: W
```

Timeseries data



The Suite in action for bare metal nodes

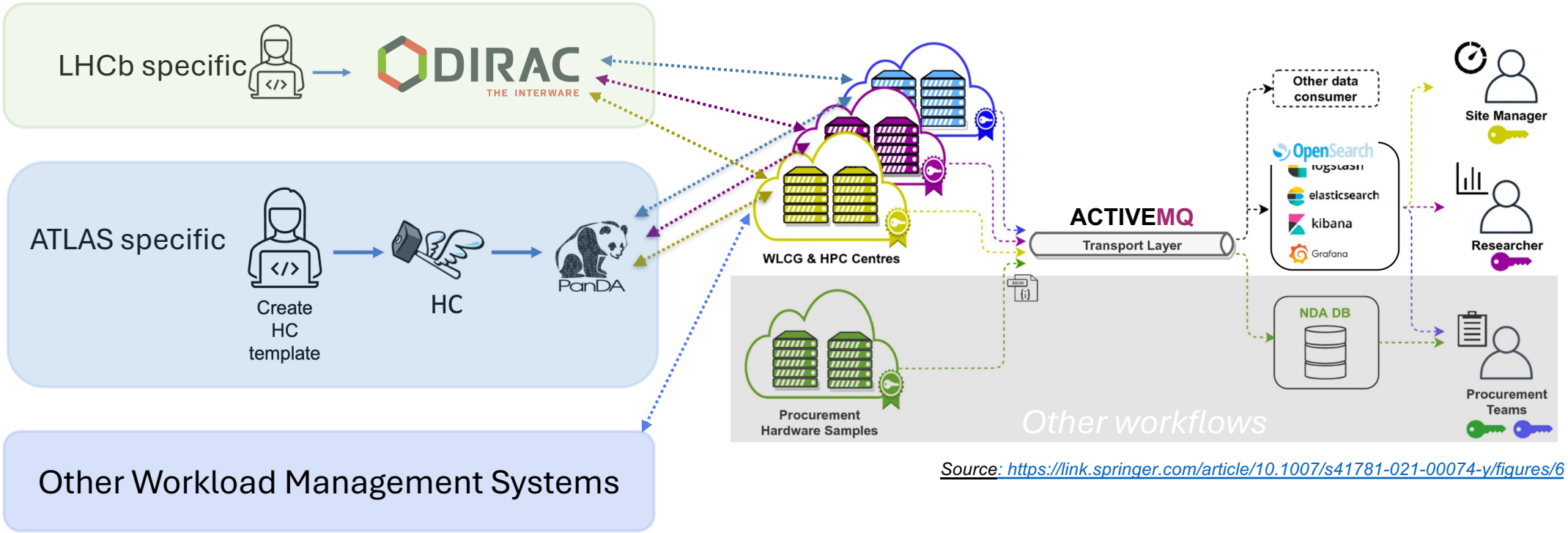
- 2 years of server benchmarks collected via the Suite and HEPScore23
- Measurements performer by sys-admins at ~26 sites
  - ~155 distinct configurations (CPU type, HT, RAM…)
  - ~14 ARM-based server configurations: Neoverse-N1, Neoverse-V2
- At CERN, all newly installed servers (>2500 so far) undergo benchmarking during acceptance
- Since release of Suite v3 (April 2025) power consumption is included in the benchmark report



Reminder

Submission Infrastructure

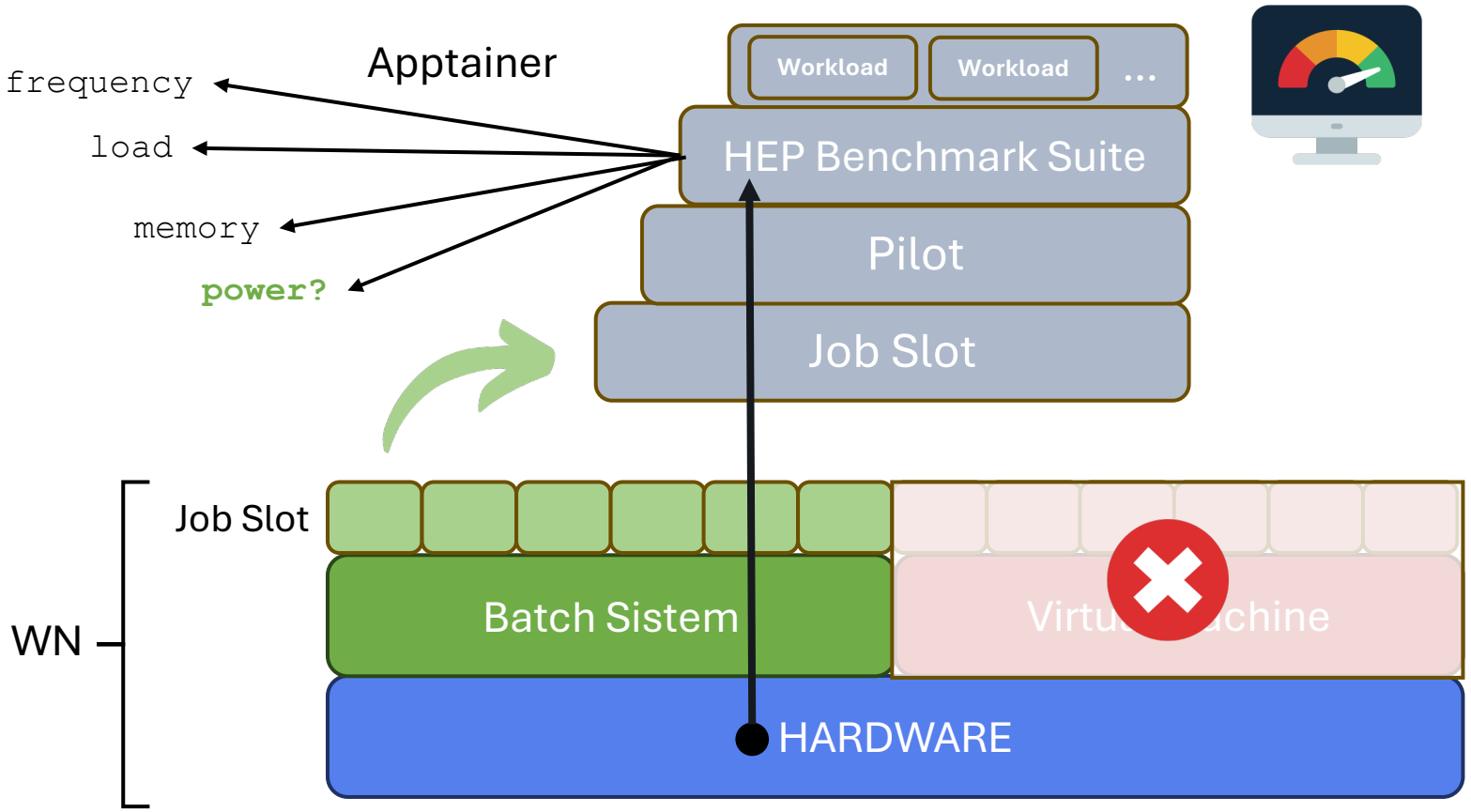
- The benchmarking script is injected in the site job slot via the standard job submission system
- Successfully implemented and deployed the pipeline for:
  - **ATLAS:** Automated submission via HammerCloud, **over 200k jobs** finished
  - **LHCb:** Manual submission to DIRAC, **2.1k** jobs finished



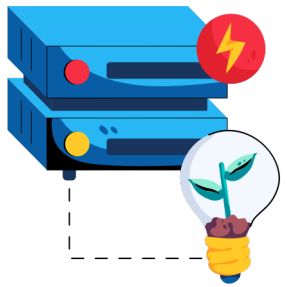
Source: https://link.springer.com/article/10.1007/s41781-021-00074-y/figures/6

Reminder

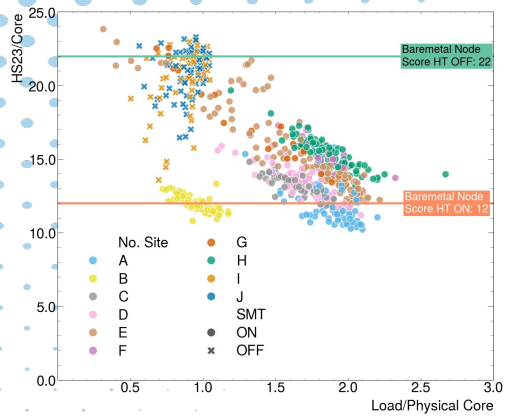
Data Collection from Pilot Jobs



Plugins run alongside the benchmark and collect information about the whole server



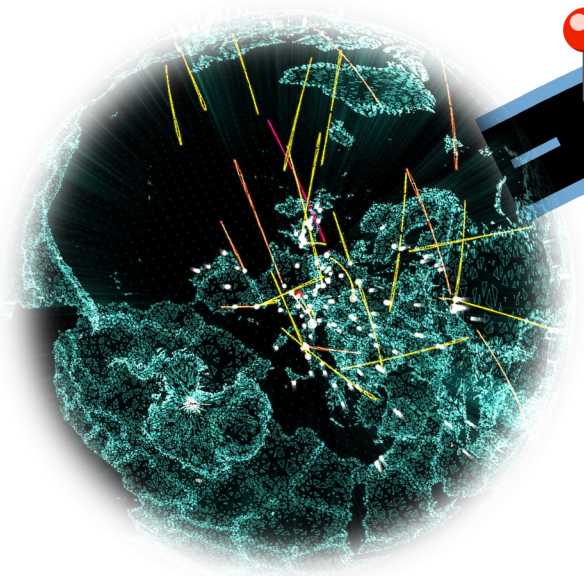
TIMELINE



1

2023: Probing Job Slots with HS23

- Automated Submission
- Collection Of Utilisation Metrics



2

2024: Data Analysis

- +200k Data Collected
- Impactful Insights
- Fixed Sites Misconfigurations
- Corepower Recalibartion

3

2024 WLCG Sustainability Workshop

- Demand of power measurements per job
- Static attempts
- Rough estimates of power consumption

4

2025: HEPiX, Lugano

- The Benchmarking WG Report
- We proposed to include and collect power measurements

5

TODAY: WLCG Workshop

- Prototype Approach, Benchmarking WG with DESY, Glasgow and AGLT2
- Automated Power Measurements

CAN WE DO BETTER?

Natalia Szczepanek (CERN)
Domenico Giordano (CERN)

Power measurements from job slots

- Can the same approach of the Suite of collecting power measurements be used in job slots probing?

- Can it be extended to other jobs, pilots?

Natalia Szczepanek (CERN)
Domenico Giordano (CERN)

Power measurements from job slots

- Can the same approach of the Suite of collecting power measurements be used in job slots probing?
  - YES

- Can it be extended to other jobs, pilots?
  - Motivation to work together and find a common solution.

Natalia Szczepanek (CERN)
Domenico Giordano (CERN)

Challenges in Grid-Level Power Consumption Measurement

1. Power consumption data is typically retrieved using system-level tools and commands that require elevated (sudo) privileges, such as **IPMI**, **Redfish**, lm-sensors, or vendor-specific interfaces...
2. It poses a security concern to make power measurement tools accessible from job slots.

Solution:

Toward a Unified Power Monitoring Approach

- There is currently no standardized approach across sites
- **We have successfully adapted to the various existing configurations, including differences in output formats, file extensions, and file locations**

DESY	• systemd service • save logs to `.current` file in `/etc/`
AGLT2	• cronjob • save logs to the `.txt` file in `/var/`
Glasgow	• cronjob • save logs to `.json` file in `/tmp/` and copy to the `/var/`

Done So Far

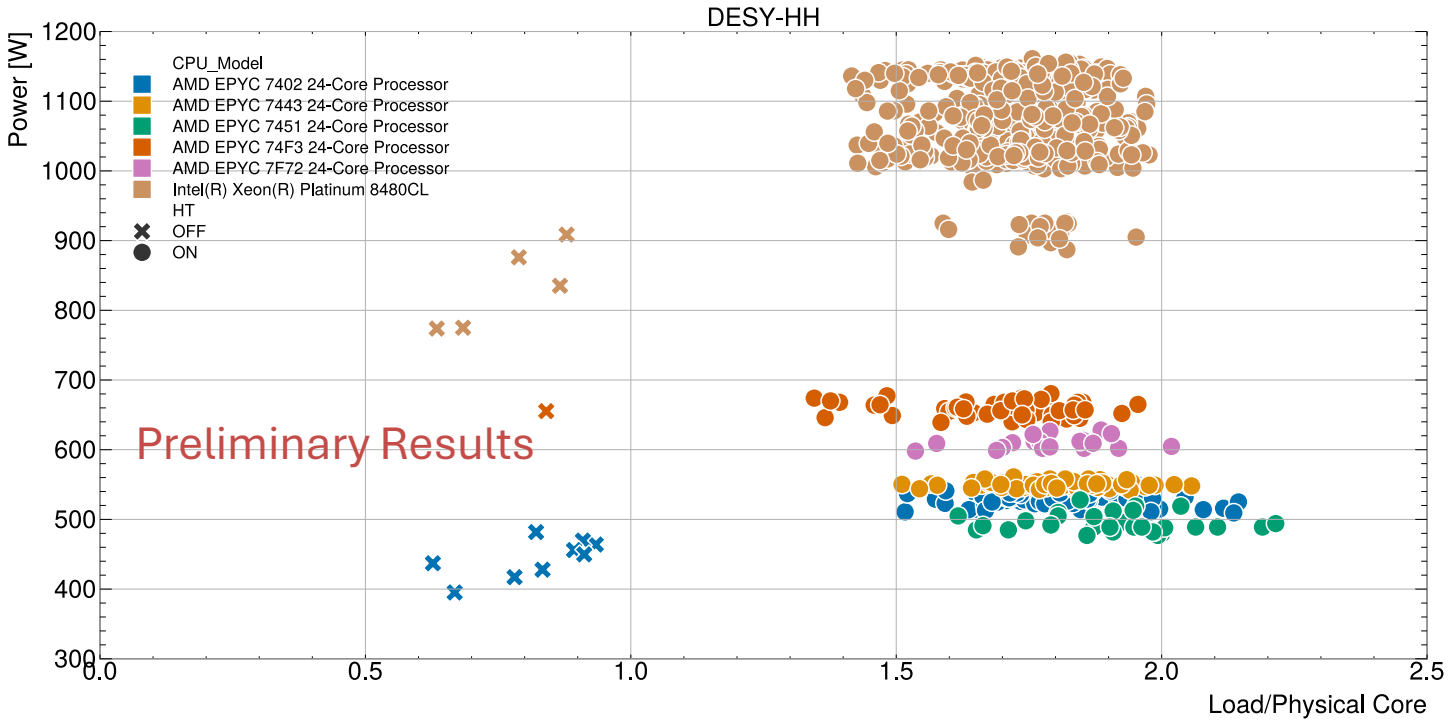
- Close collaboration with DESY, AGLT2 and Glasgow
- Submitted and collected data from +2000 jobs in the last 2 weeks
- Probed 21 different CPU Models on 3 sites
- Preliminary data in the next slides
  - Next step: data validation, cleaning, calibration

Natalia Szczepanek (CERN)
Domenico Giordano (CERN)

DESY-HH

```
power-consumption:
  command: "cat /etc/batch/power-consumption/power.current | cut -d' ' -f2"
  regex: '(?P<value>\d+)'
  interval_mins: 1
  unit: 'W'
```

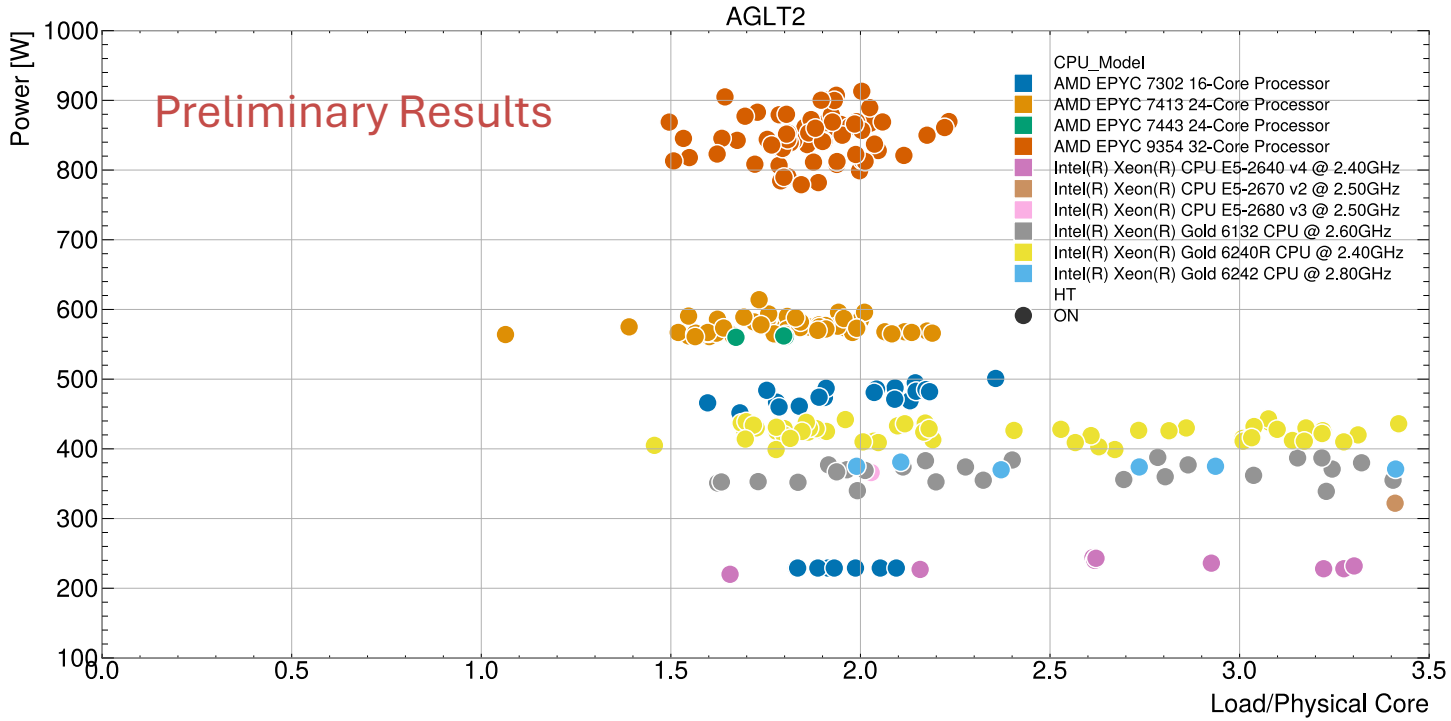
CPU Model	HT	Power [W]	Count
AMD EPYC 7402 24-Core Processor	0	437.05	7
AMD EPYC 7402 24-Core Processor	1	531.175	84
AMD EPYC 7443 24-Core Processor	1	549.9	79
AMD EPYC 7451 24-Core Processor	1	489.0	37
AMD EPYC 74F3 24-Core Processor	0	655.3	1
AMD EPYC 74F3 24-Core Processor	1	659.5	52
AMD EPYC 7F72 24-Core Processor	1	604.7	19
Intel(R) Xeon(R) Platinum 8480CL	0	825.5	4
Intel(R) Xeon(R) Platinum 8480CL	1	1072.45	529
AMD EPYC 7402 24-Core Processor	0	437.05	7
AMD EPYC 7402 24-Core Processor	1	531.175	84



AGLT2

```
power-consumption:
  command: 'cat /var/log/power-reading.txt'
  regex: 'Instantaneous power reading:\s*(?P<value>\d+) Watts'
  interval_mins: 1
  unit: 'W'
```

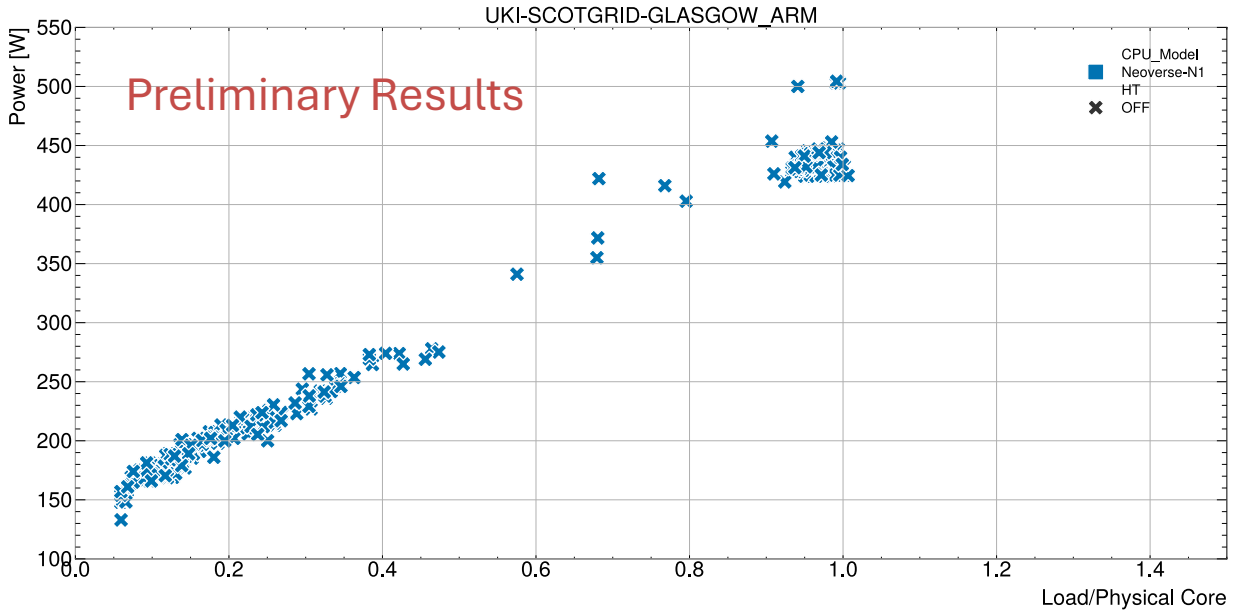
CPU Model	HT	Power [W]	Count
AMD EPYC 7302 16-Core Processor	1	465.0	47
AMD EPYC 7413 24-Core Processor	1	574.325	109
AMD EPYC 7443 24-Core Processor	1	561.0	4
AMD EPYC 9354 32-Core Processor	1	857.3	66
Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40GHz	1	233.0	13
Intel(R) Xeon(R) CPU E5-2670 v2 @ 2.50GHz	1	333.0	2
Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz	1	366.0	5
Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz	1	362.0	39
Intel(R) Xeon(R) Gold 6240R CPU @ 2.40GHz	1	425.5	84
Intel(R) Xeon(R) Gold 6242 CPU @ 2.80GHz	1	375.0	6

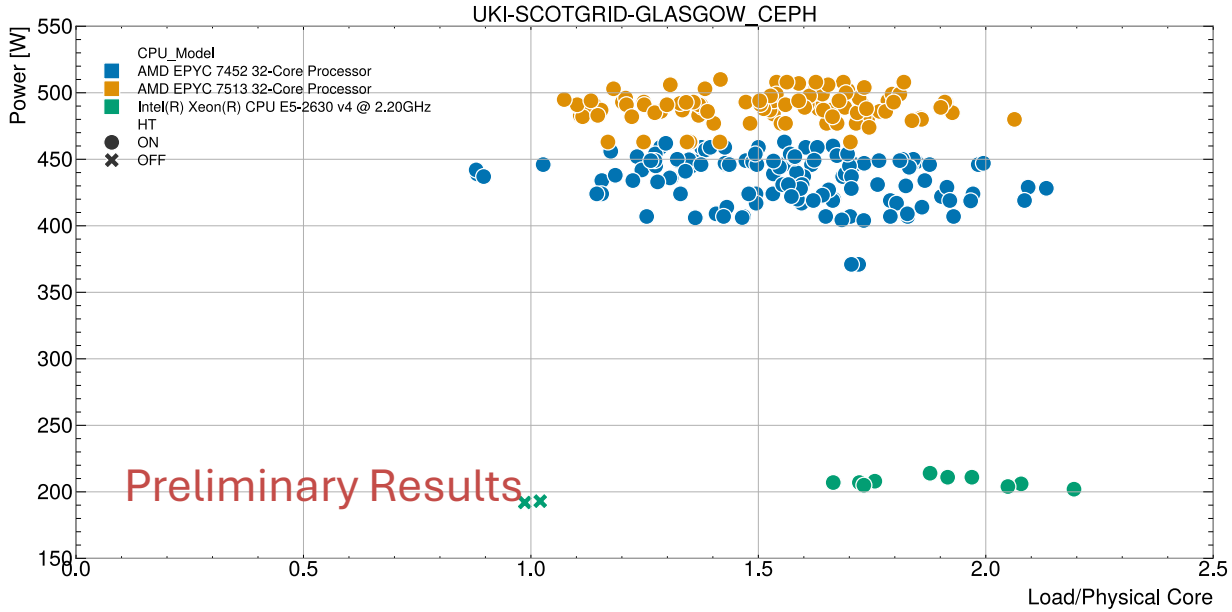


GLASGOW

CPU Model	HT	Power [W]	Count
Neoverse-N1: Ampere Altra Max M128-30	0	213.3	207
AMD EPYC 7452 32-Core Processor	1	428.0	230
AMD EPYC 7513 32-Core Processor	1	488.0	122
Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz	0	192.5	2
Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz	1	211.0	9

```
power-consumption:
  command: 'cat /var/log/power-reading.txt'
  regex: 'Instantaneous power reading:\s*(?P<value>\d+) Watts'
  interval_mins: 1
  unit: 'W'
```





Summary and Next Steps

Contact us, subscribe to the HEPiX Benchmarkign WG list https://listserv.in2p3.fr/cgi-bin/wa?SUBED1=hepix-cpu-benchmark

Proposed strategy: expose node-level power metrics in jobs (or pilots) to enrich job monitoring reports

Feasibility demonstrated through collaboration with DESY, AGLT2, and Glasgow

Identify **a unified approach** for the metric availability at sites

- Common script/daemon with standardized data format and file placement
- Ensure accessibility from job slots and compatibility with pilot frameworks

Scale site adoption

- Pro-tempore using ATLAS HammerCloud tests to onboard new sites.
- Should become an ETF (old SAM) test.

Coordinate and collaborate on a central data collection and validation

The Unified Experiment Monitoring is an excellent candidate for that

Take part to the effort of the HEPiX Benchmarking WG,

Engage the wider WLCG community to align efforts and promote sustainability together

Appendix

Isolated Tests: Power Measurements on Baremetal Nodes