Towards Unified Power and Efficiency Monitoring Across the Worldwide LHC Computing Grid

Natalia Szczepanek,
Domenico Giordano
CERN, Geneva

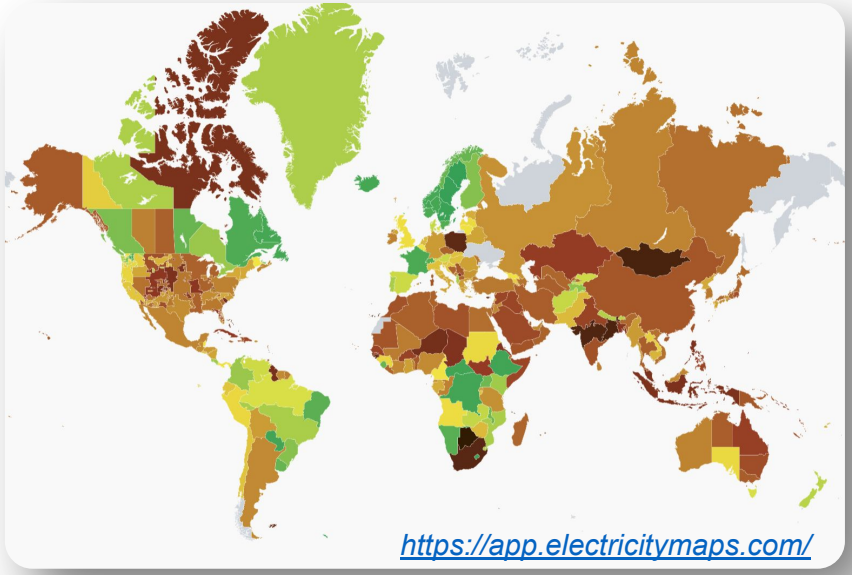
Sustainable HEP 2026 - 5th edition

Importance of sustainability in WLCG

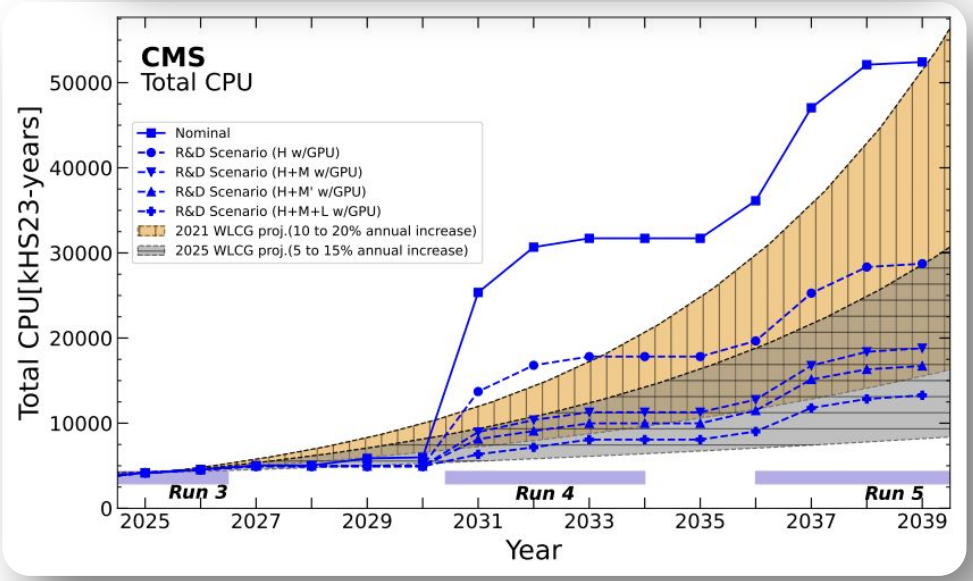
~200 Sites
across 42 countries

80%
of computing power sits outside CERN

~1.4M CPU cores
used across WLCG



- HL-LHC Computing Needs:
  - Pessimistic scenario: x5 higher than in Run-3
  - Optimistic scenario: 40% higher than in Run-3
- Environmental impact of computing is strictly connected to **financial savings**



https://cds.cern.ch/record/2957472/files/LHCC-G-186.pdf

A better understanding of power consumption in WLCG will allow to make better, data-driven decisions.

Where are we? Three pillars of WLCG strategy

WLCG Strategy 2024–2027 – three environmental recommendations guiding the community

ENV-1

Measure & Monitor

Agree on metrics and a framework to collect energy-efficiency data.

Where we are:

- Ensure Project enabling power-consumption data collection
- WLCG starting to collect PUE per site

ENV-2

Efficient hardware

Facilitate more energy-efficient hardware as experiment software matures.

Where we are:

- GPU-capable generators, simulation & reconstruction
- Task Force discussing how to pledge GPU resources

ENV-3

Sustainability plan

Cut carbon footprint across software, computing models, facilities & hardware lifecycle.

Where we are:

- Longer-term – defining scope for a global collaboration
- Early work: heat re-use, ARM/GPU tech, optimal HW lifetime

Guided by the four 'M's – **Measure, Model, Monitor, Moderate**.
What can we measure and how? Model the system to understand optimizations.
Develop monitoring of performance. Moderate usage where possible.

1 The problem

No standardized power monitoring across WLCG sites

10 W/core - over or underestimation?

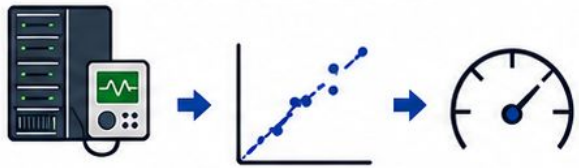
Vendor TDP ≠ real power under HEP workloads

PUE unknown per site

No common method to estimate carbon from grid jobs

2 Methodology

Power measurements



Collect (Power + HS23 + other metrics) → Model → Predict

TDP analysis

TDP vendor specs + Power measurements + WLCG job statistics

3 Results

Load-aware power prediction model

Site-level anomaly detection

Derived estimation for power/core

Extensible power database for the community (Estimated W/Core per site)

Extend grid studies by:

HS23/W metric for hardware procurement

4 Impact

First empirical bottom-up approach for WLCG-scale energy characterization

Provides the data foundation for energy-aware hardware procurement decisions

Outcome

A data-driven foundation for accurate power estimates, informed energy planning, and transparent carbon accounting across WLCG.

Better accuracy | Energy awareness | Informed decisions | Community alignment

Power Measurements

HEP Benchmark Suite

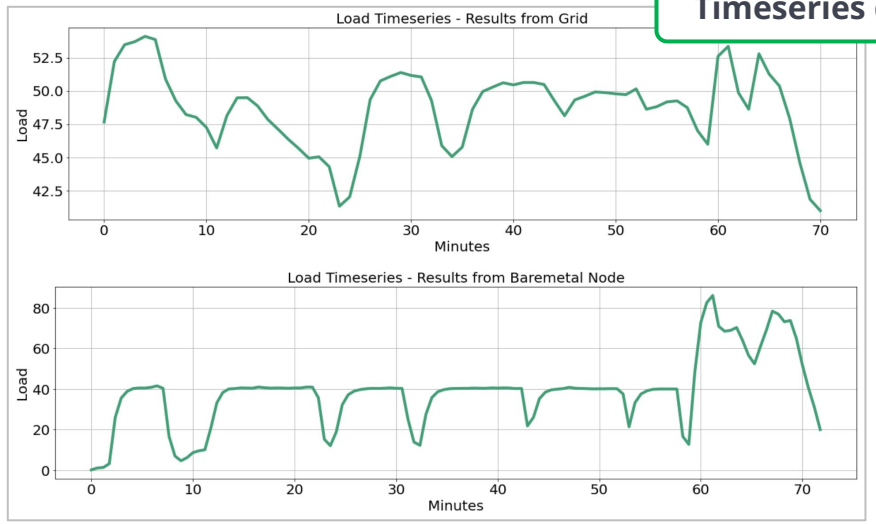
hep-benchmark-suite

- 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

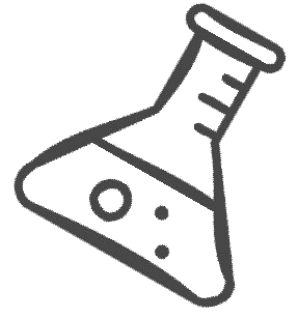
```
plugins:
  CommandExecutor:
    metrics:
      cpu-frequency:
        command: >-
          cpupower -c all frequency-info -f | grep 'current CPU frequency:' |
          grep -o '[0-9]\{7,\}' | awk '{s+=$1; c++} END {print (s/c)/1000}'
        interval_mins: 1
        regex: (?P<value>\d+.\d+).*
        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



Measuring Performance Under Diverse Conditions

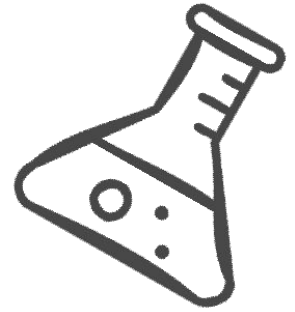
- **Previous examples mainly covered isolated, controlled conditions: *the pledge use case***
  - Server executes **exclusively** the benchmark application
  - Ensure predictable behaviour and consistent measurement conditions



- **Real-world conditions: *the batch job use case***
  - Thanks to the **free license**, the benchmark can be executed over the Grid
  - Running with **limited cores via Grid job submission**, while the rest of the server runs diverse, unpredictable **production** applications
  - Allows sampling of virtually all grid nodes under **practical workload scenarios**

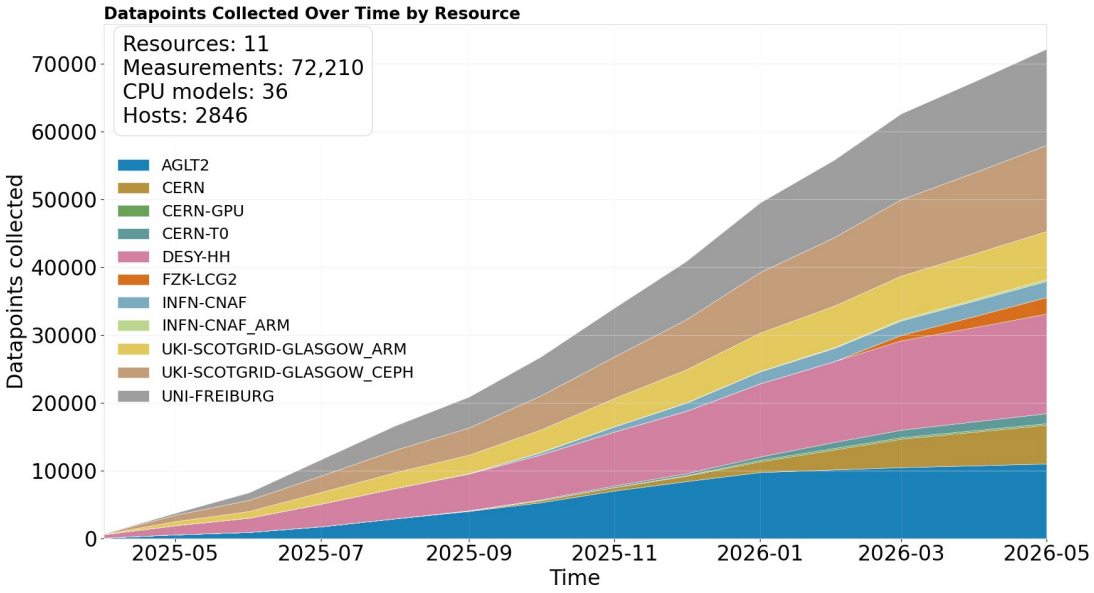
Measuring Performance Under Diverse Conditions

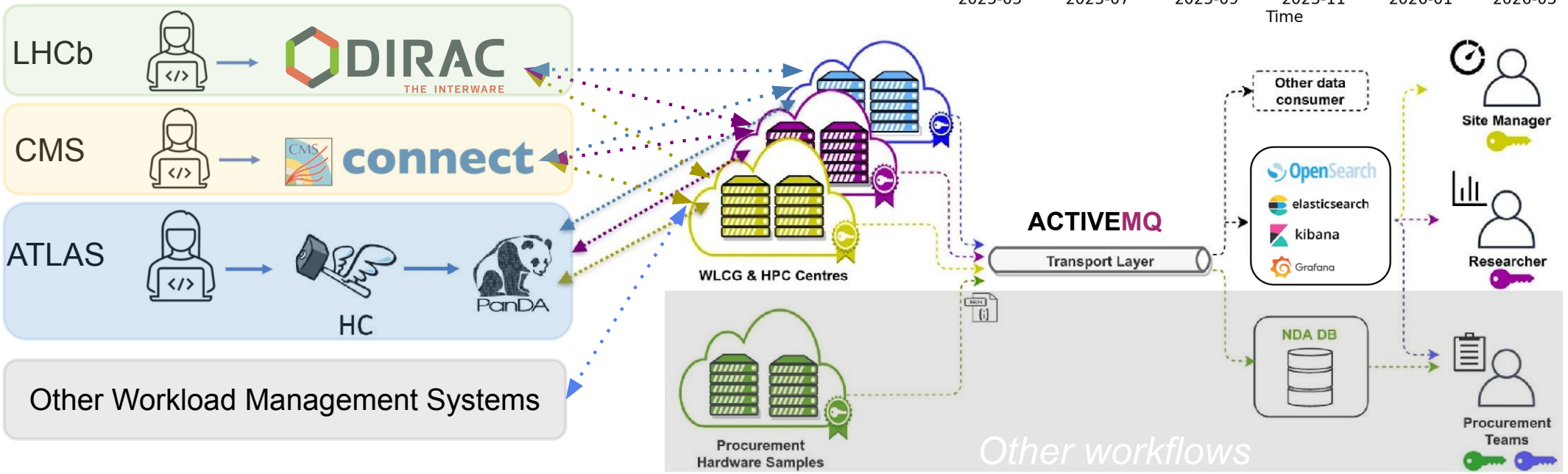
- **Previous examples mainly covered isolated, controlled conditions: *the pledge use case***
  - Server executes **exclusively** the benchmark application
  - Ensure predictable behaviour and consistent measurement conditions



- **Real-world conditions: *the batch job use case***
  - Thanks to the **free license**, the benchmark can be executed over the Grid
  - Running with **limited cores via Grid job submission**, while the rest of the server runs diverse, unpredictable **production** applications
  - Allows sampling of virtually all grid nodes under **practical workload scenarios**

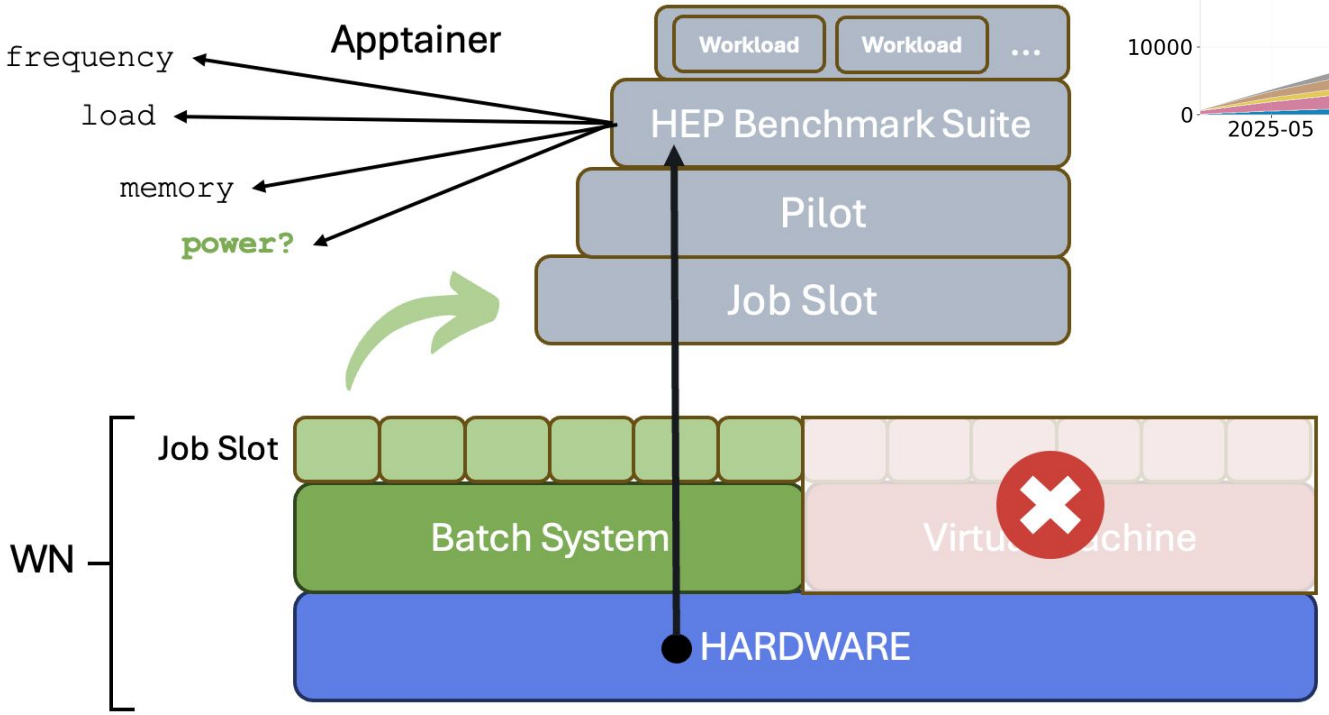
Leverage existing submission infrastructure

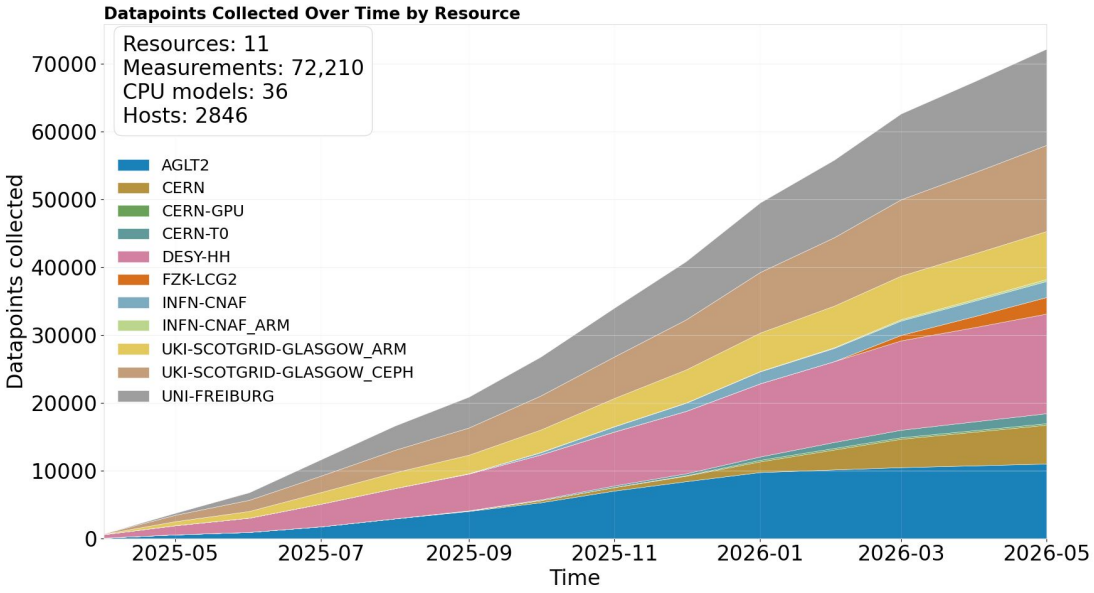




Source: https://link.springer.com/article/10.1007/s41781-021-00074-y/figures/6

Leverage existing submission infrastructure



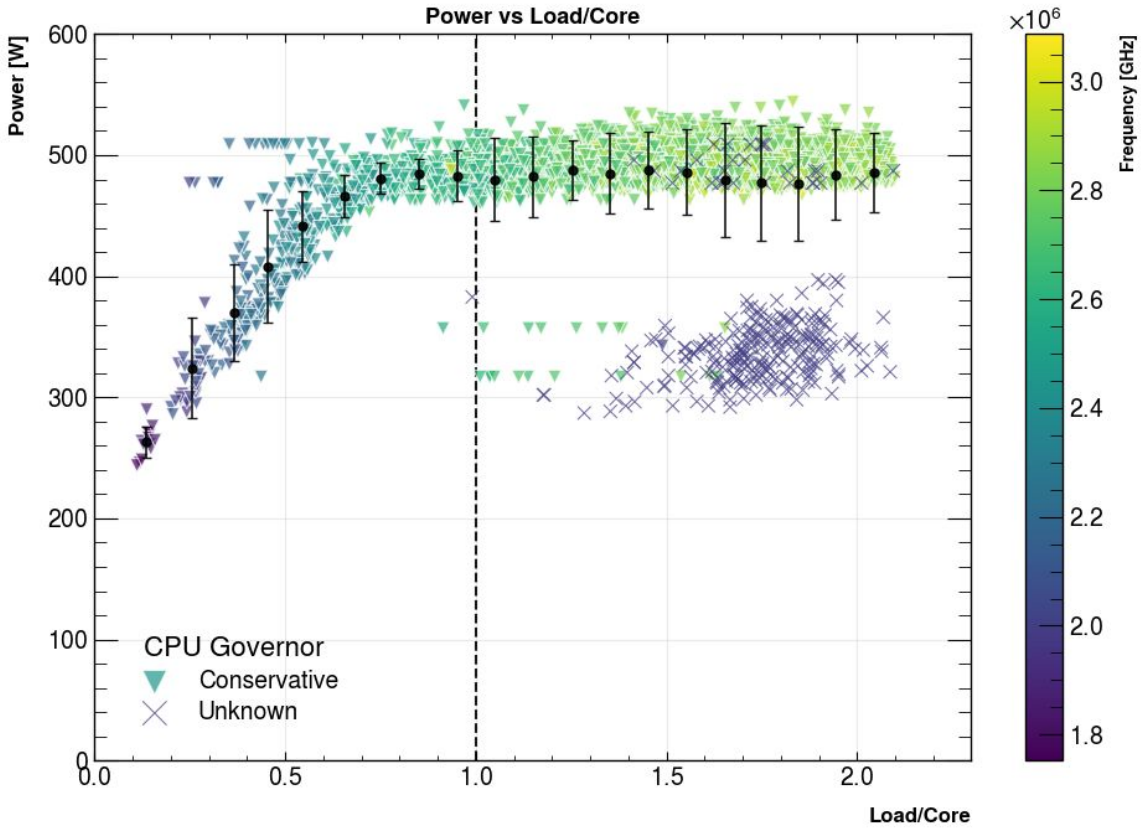


Plugins run alongside the benchmark and collect information about the whole server

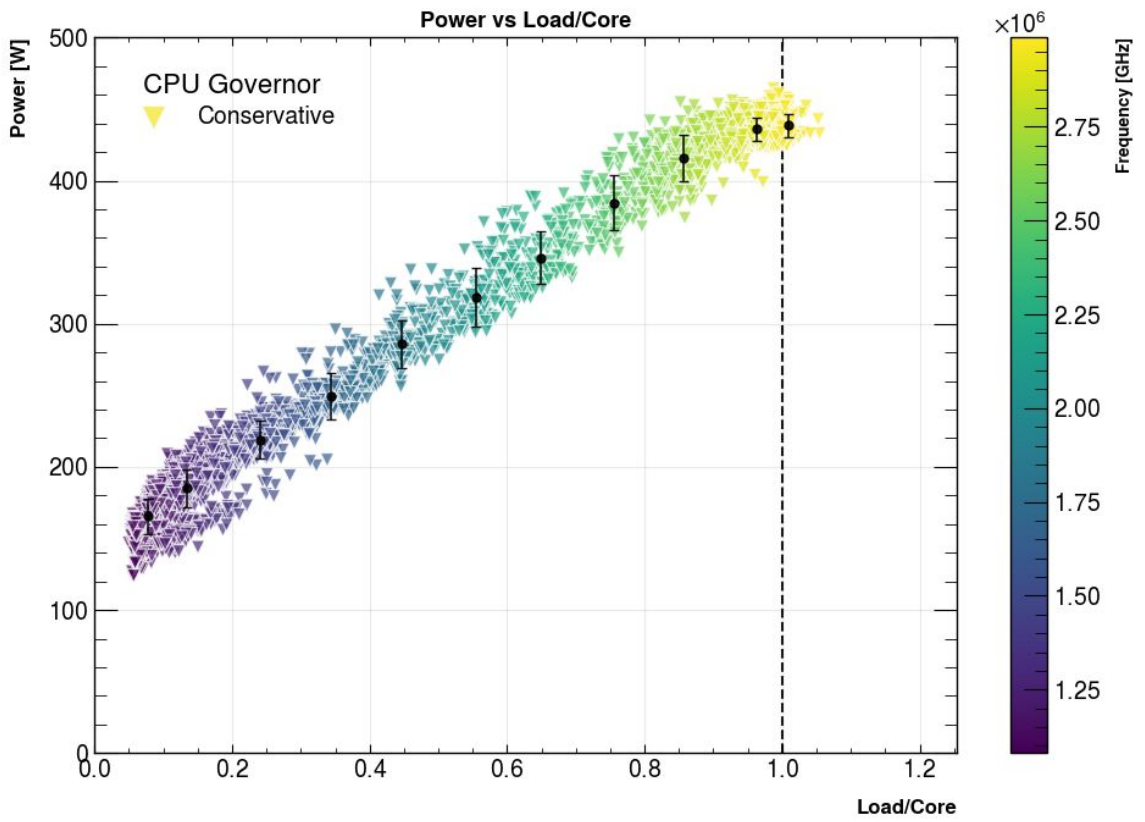
https://gitlab.cern.ch/hep-benchmarks/hep-power-collector

Collected measurements from Grid nodes

AMD EPYC 7513 32-Core



Neoverse-N1



Power shows a saturating trend with load ~ 1 (x86 - HT enabled) in load range (0-2)
We need to differentiate noise and outliers from deliberate server configurations

Looking ahead: Dynamic Power Prediction

- We intend to provide the community with a dynamic power prediction tool
- **Proof of concept**, to be fully developed
- The goal is to collect as much reliable fitted models as possible to cover WLCG needs
- Given a CPU model and load, the tool can predicts idle, current and full-load power without direct measurements
  - An accurate validation on a test dataset is needed

TDP and Power Analysis

* TDP - Thermal Design Power

Strategy

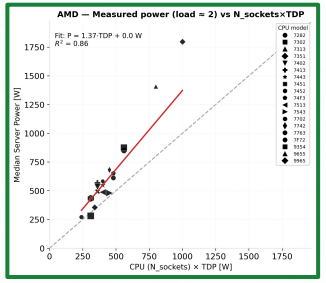
1 Collect CPU TDP data

Collect CPU TDP data from official vendor specifications for the broadest possible coverage of deployed CPU models.

Model	Architecture	TDP (W)
Xeon 8480+	Sapphire Rapids	350
EPYC 9654	Genoa	360
EPYC 9554	Genoa	360
...	...	...

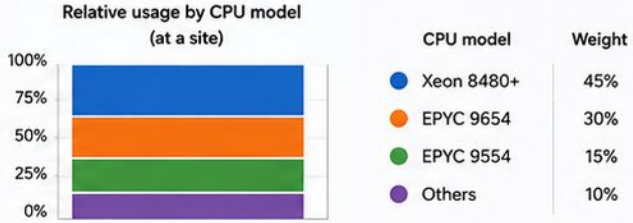
2 Derive scaling factors

Use real power measurements from previous step to derive architecture-specific scaling factors that translate TDP into estimated server power.



3 Determine usage weights

Use 2025 production job statistics, based on walltime × cores, to determine the relative usage weight of each CPU model at every site.



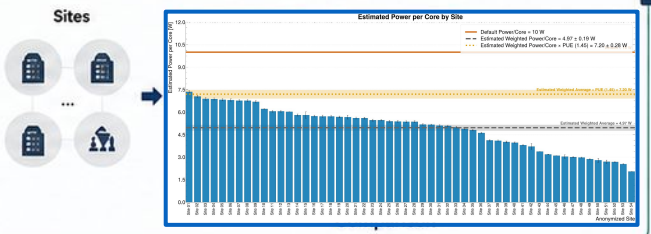
4 Compute weighted power/core per site

Compute the weighted estimated power per core for each site using the derived scaling factors and site-specific CPU fleet composition.

Site CPU mix (weights)		×	Scaling factor (Server Power / TDP)		=	Estimated power/core (W/core)
Xeon 8480+	45%		Xeon (SPR)	1.45		X
EPYC 9654	30%		EPYC 9654	1.40		
EPYC 9554	15%		EPYC 9554	1.38		
Others	10%		Others	...		

5 Derive WLCG-wide power/core and compare

Aggregate weighted site-level values to obtain a WLCG-wide power/core estimate, then apply an assumed average PUE to compare against the current default assumption of 10 W/core.



6 Develop an automated calculator

Develop an automated calculator to estimate site power consumption from fleet composition.

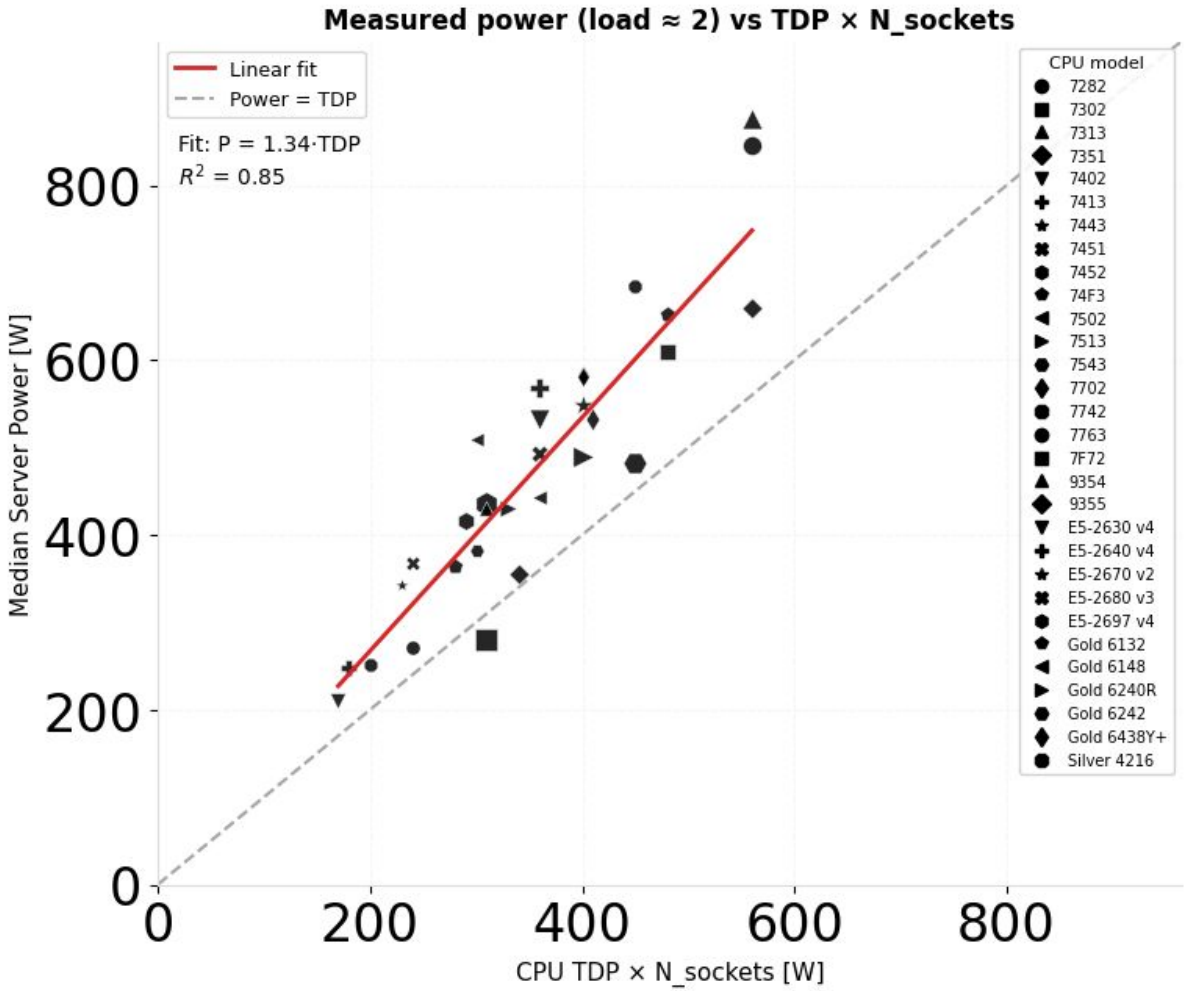
Outcome

Data-driven, site-specific and WLCG-wide power estimates for informed energy planning and realistic comparisons.

- Better accuracy
- Energy awareness
- Informed decisions
- Community alignment

CPU TDP vs server IPMI

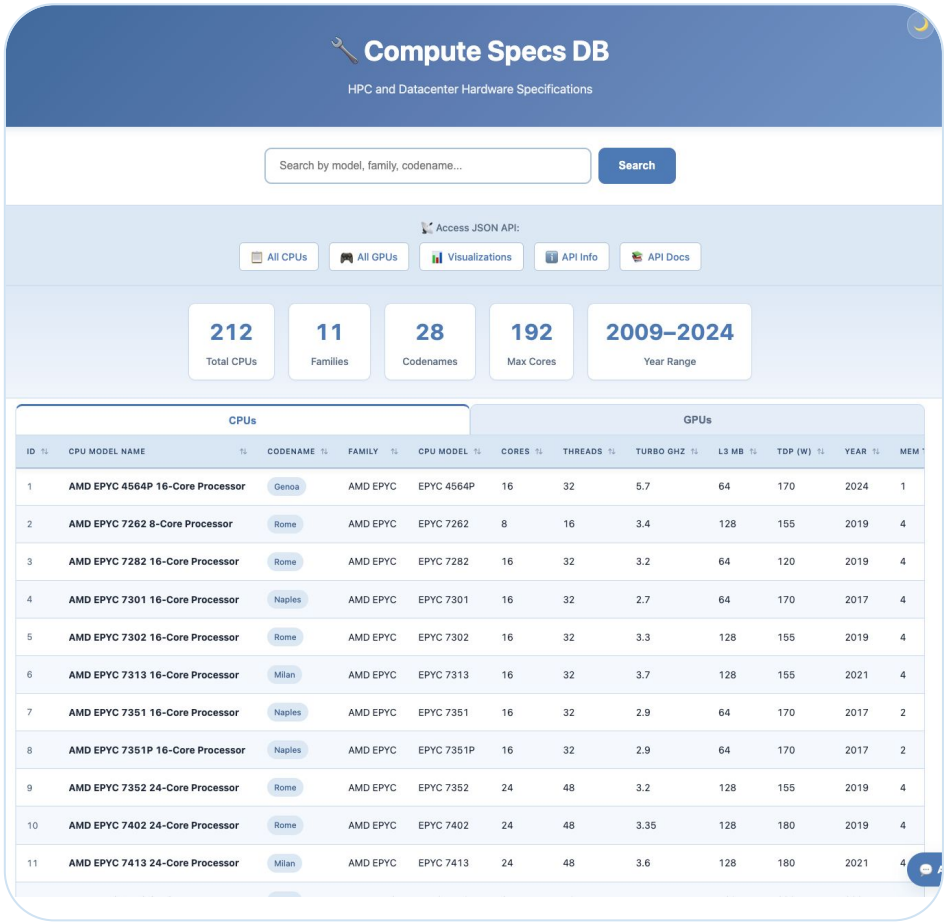
- TDP stands for Thermal Design Power, in watts, and refers to the power consumption under the maximum theoretical load.
- TDP applies only to the chip itself (CPU or GPU), not the whole system.
- Measured power at fully loaded machines vs TDP shows linear relation that could be used if power measurements are not possible

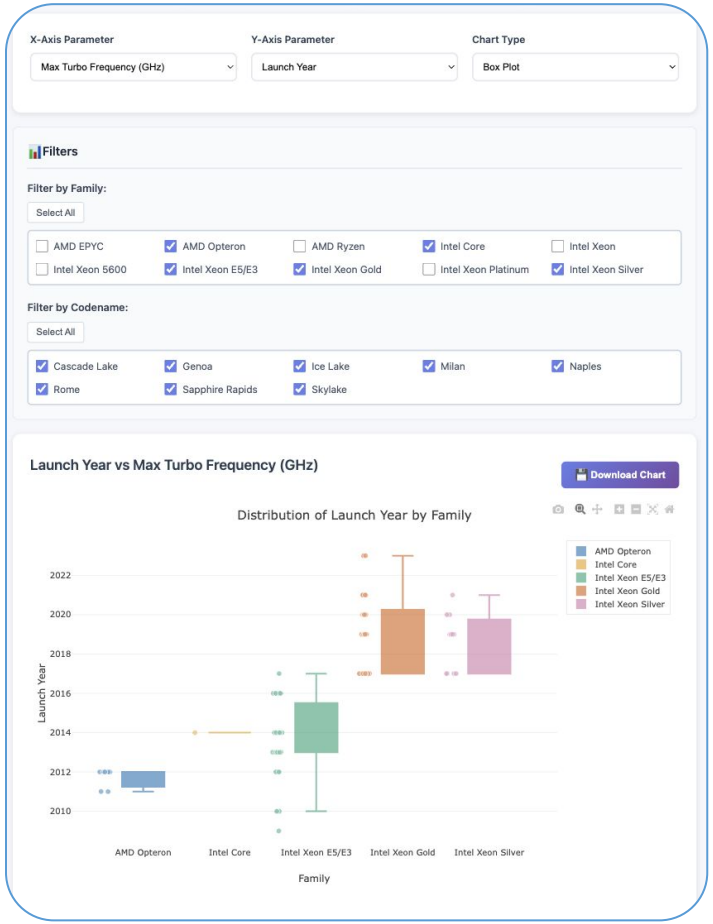


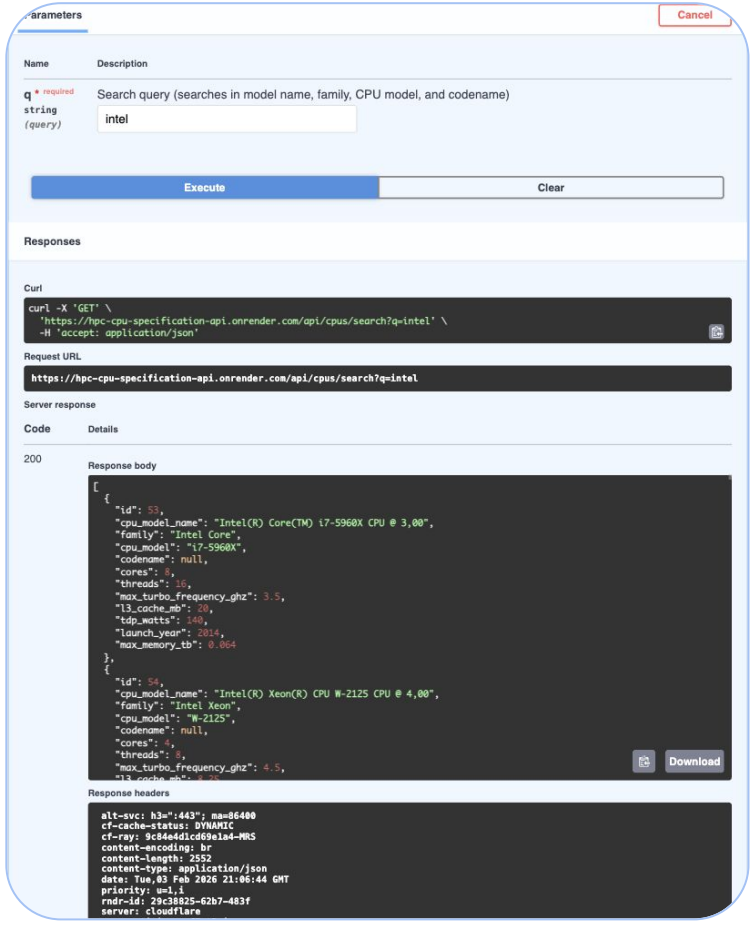
CPU TDP dataset

- Collected 212 CPU model specs from official data sources
- Tables with CPU and GPU info, visualizations, API; **very useful for analysis**
- PoC, plan to be extended in the future

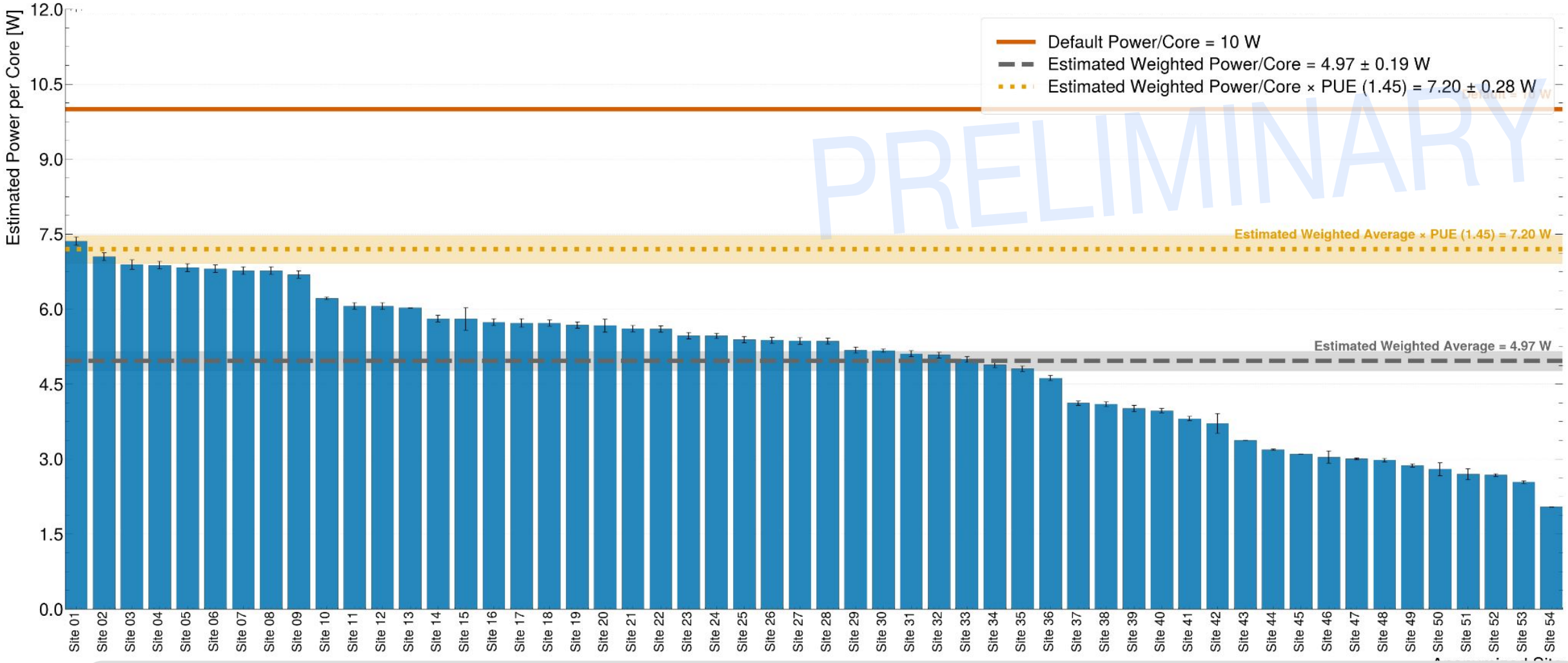
https://computespecsdb.com/







What is the estimated power per core of some WLCG sites?



Empirical bottom-up approach to estimate power/core per site and WLCG Average:
This is a subset of WLCG Sites, using available data from ATLAS → will be extended to entire WLCG
Estimated Weighted Average Power/Core scaled by PUE (1.45) is 7.2W/core

Goal:
Develop a "Site Energy Calculator"

🏢 Site Energy Calculator

TDP-based fleet estimation · AMD ×1.37 Intel ×1.50 rescale factors · HEPiX / WLCG

FLEET CONFIGURATION

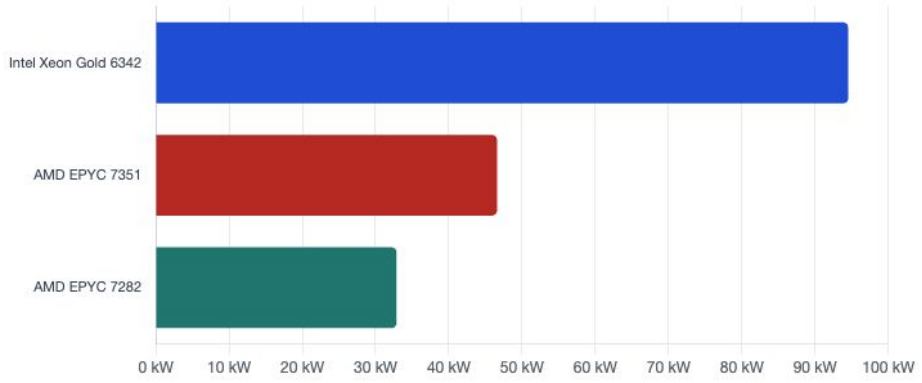
Clear all | + Add CPU row

CPU MODEL	NODES	SOCKETS / NODE	W / SOCKET	TOTAL CORES	TOTAL POWER
AMD EPYC 7282 (16c · 164.4 W)	100	2	164 W	3,200	32.88 kW
AMD EPYC 7351 (16c · 232.9 W)	100	2	233 W	3,200	46.58 kW
Intel Xeon Gold 6342 (24c · 345 W)	137	2	345 W	6,576	94.53 kW

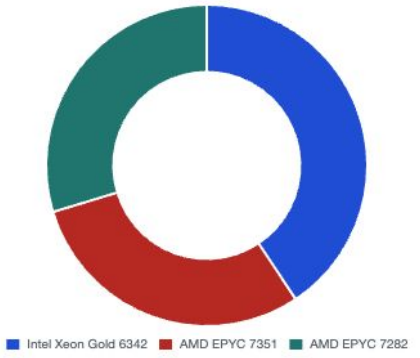
PUE 1.4 Power Usage Effectiveness = total facility power ÷ IT power (1.35 = CERN typical)

TOTAL NODES	TOTAL SOCKETS	TOTAL CORES	AVG W / CORE	IT POWER (NO PUE)	FACILITY POWER (+ PUE)
337 nodes	674 CPU sockets	12,976 physical cores	13.4 W W per physical core	173.99 kW 1524.2 MWh / year	243.59 kW 2133.8 MWh / year

POWER CONTRIBUTION BY CPU MODEL (KW)



NODE COMPOSITION



Conclusions

What we built
(Achievements)

Leverage existing benchmarking and job submission infrastructure to collect power measurements together with performance and other metrics

Load-aware power model
$P = a \cdot (1 - e^{-b \cdot L}) + c$
fitted to real power measurements

Vendor TDP-based pipeline enabling site-level power estimates without direct measurements

Key result

Estimated weighted average of power/core

≈ 7.2 W/core

(with PUE 1.45)*

*https://www.epj-conferences.org/articles/epjconf/pdf/2024/05/epjconf_chep2024_04001.pdf

Next steps

Extend load-aware power model coverage and validate predictions

Automated reports for sites

Extensible power database for the community

Move the study to HS23/Watt for hardware procurement

Impact

Stronger data foundation for energy-aware decisions across the WLCG community.

Better accuracy

Energy awareness

Informed decisions

Community alignment