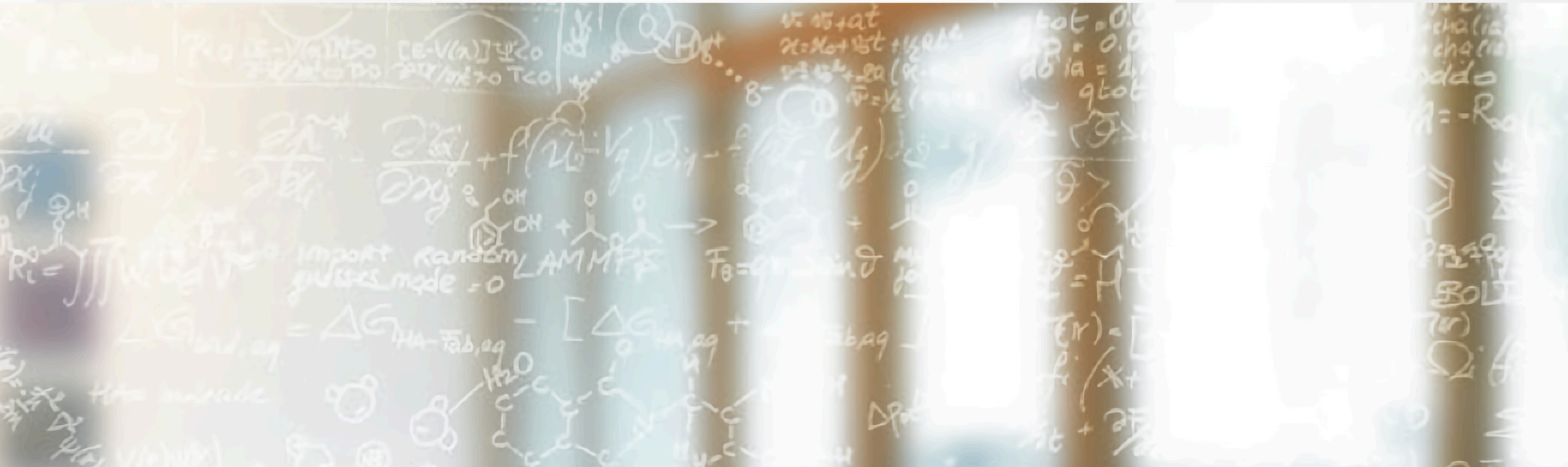
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



EMOI: CSCS Extensible Monitoring and Observability Infrastructure

Jean-Guillaume Piccinali, May 16th 2024

5^{th} ISC HPC MODA24 Workshop on Monitoring & Operational Data Analytics

https://moda.dmi.unibas.ch

Outline

- Motivation
- EMOI: CSCS Extensible Monitoring and Observability Infrastructure
- Use case: Power measurements and Energy dataset

Disclaimer

- New working group and new (pre-acceptance) system
- This talk is not about power saving techniques (yet)

Motivation

Sustained Performance and Power Efficiency

Top500, Green500	Top 6 Systems (June 2024)	Power Eff. Green500	Power Top500
#1, #7	Frontier/ORNL	62.684 GF/W	22.786 kW
#2, #42	Aurora/ANL	26.151 GF/W	38.698 kW
#3, #xx	Eagle/Microsoft Azure	-	-
#4, #68	Fugaku/RIKEN	15.418 GF/W	29.899 kW
#5, #12	LUMI/EuroHPC	53.428 GF/W	7.107 kW
#6, #14	Alps/CSCS	51.983 GF/W	5.194 kW

Green500 #1 Power Efficiency

Green500	Efficiency
2019/06	15.1 GF/W
2019/11	16.9 GF/W
2020/06	21.1 GF/W
2020/11	26.2 GF/W
2021/06	29.7 GF/W
2021/11	39.4 GF/W
2022/06	62.7 GF/W
2022/11	65.1 GF/W
2023/06	65.4 GF/W
2023/11	65.4 GF/W
2024/06	72.7 GF/W

Idle Compute Node Power Usage

Node Type	Idle Power
Intel Broadwell	183 kWh
AMD Rome	691 kWh
AMD Milan	1101 kWh
Intel Haswell + 1 NVIDIA P100	273 kWh (1 GPU = ~1/3)
AMD Milan + 4 NVIDIA A100	1951 kWh (1 GPU ~1/8)

- Left: Better and better energy efficiency in top systems / significant amount of power (22 MW for Frontier),
- Middle: 4x more energy efficient in 4 years / slowdown since November 2022 / best in June 2024,
- Right: Idle parts of a node are getting more and more energy intensive,
- Increased electricity costs in Europe since 2023

Monitoring power and energy is critical

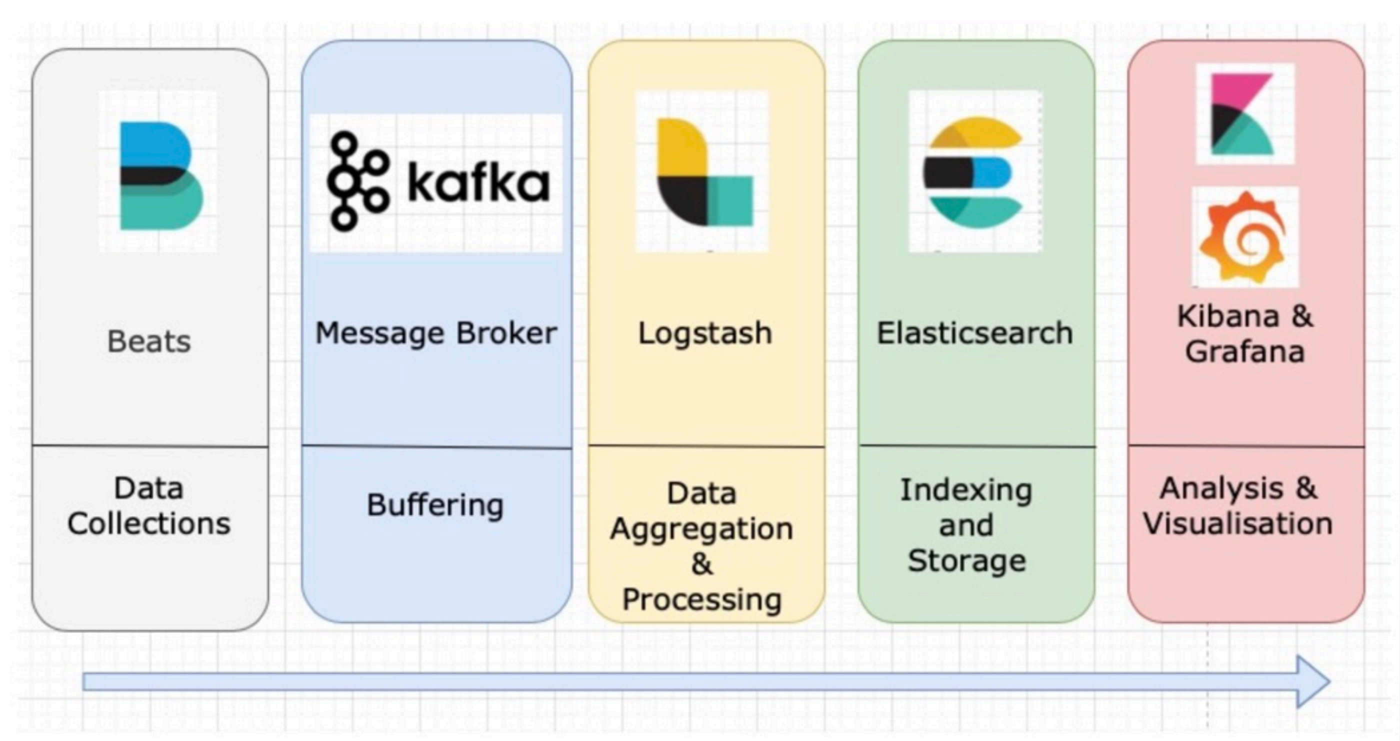
Motivation

- **Piz Daint**: NVIDIA *P100* Cray XC production sytem since 2016

- **Alps**: new multitenant heterogeneous HPE/Cray EX system
  - 2020/Phase 0: AMD Rome (zen2) CPU nodes,
  - 2022/Phase 1: NVIDIA *A100* and AMD *MI250x* GPU nodes, AMD Milan (zen3) CPU nodes,
  - 2024/Phase 2: NVIDIA Grace CPU and Hopper GPU *GH200* nodes.

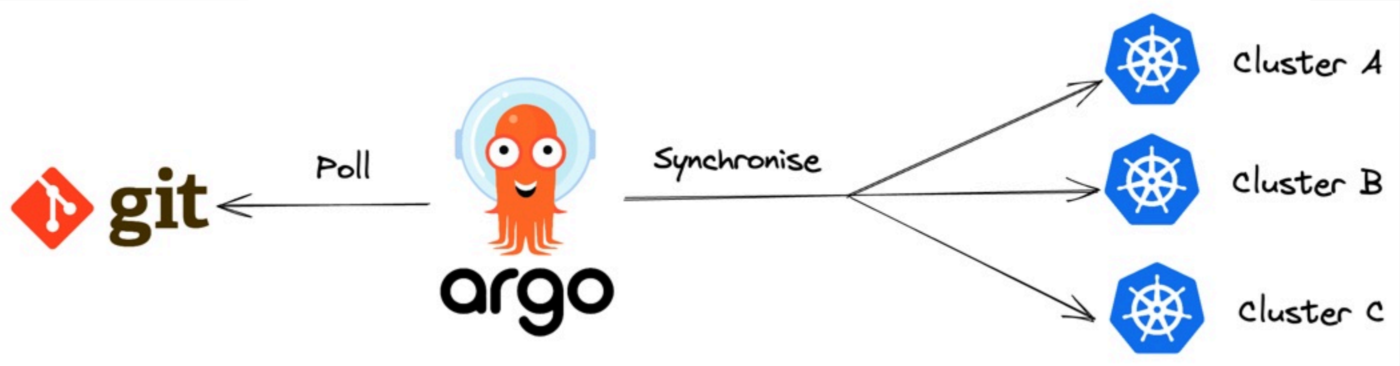
EMOI: Extensible Monitoring and Observability Infrastructure

EMOI Infrastructure components: Elastic Stack (ELK)



- **Beats**: data collection with lightweight shippers, hundreds of GB per day,
- **Kafka**: buffer and message broker, push model, streaming telemetry,
  - Integrated with HPE CSM/SMA Kafka Bus
- **Logstash**: data transformation for ES (smaller messages) and **Memcached** for data enrichment,
- **Elasticsearch**: distributed search and analytics engine designed for storing large volumes of data,
- **Kibana/Grafana**: analytics and dashboards

EMOI Infrastructure components: Elastic Cloud on Kubernetes (ECK)



- **ArgoCD**: continuous deployment of the ELK on Kubernetes,
- Benefits of a **GitOps** approach:
  - *Agility*: rapid response to changing workload demands,
  - *Efficiency*: optimized resource utilization increase operational efficiency,
  - *Stability*: configuration change tracking improve operational stability,
  - *Automation*: Infrastructure as Code allows continuous delivery of updates and new features,
- Cluster management: TerraForm, Rancher and Harvester.

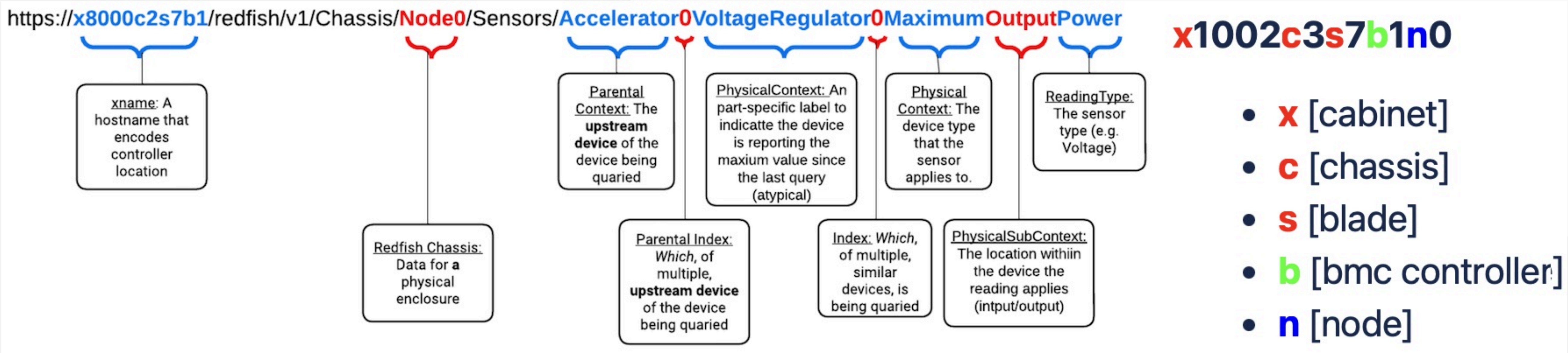
Power measurements and Energy dataset

Collecting Energy Data

- PM data: Consumed Energy at Node, CPU and GPU levels can be read from **/sys/cray/pm_counters/** sysfs files. Default collection rate is **10** Hz. The energy usage at Node level can also be accessed with the **Slurm** sacct command.

```
read pm_counters/energy when the job starts: E_t0=669376366 J # 1710250886297894 us
read pm_counters/energy when the job ends:   E_t1=669935671 J # 1710251151444267 us
get node energy of the job:                  E = E_t1 - E_t0 [J]
```

- TM data: HPE/Cray sensors are published via the **Redfish** restful API, using the Sensor schema. Default collection rate is **1** Hz.

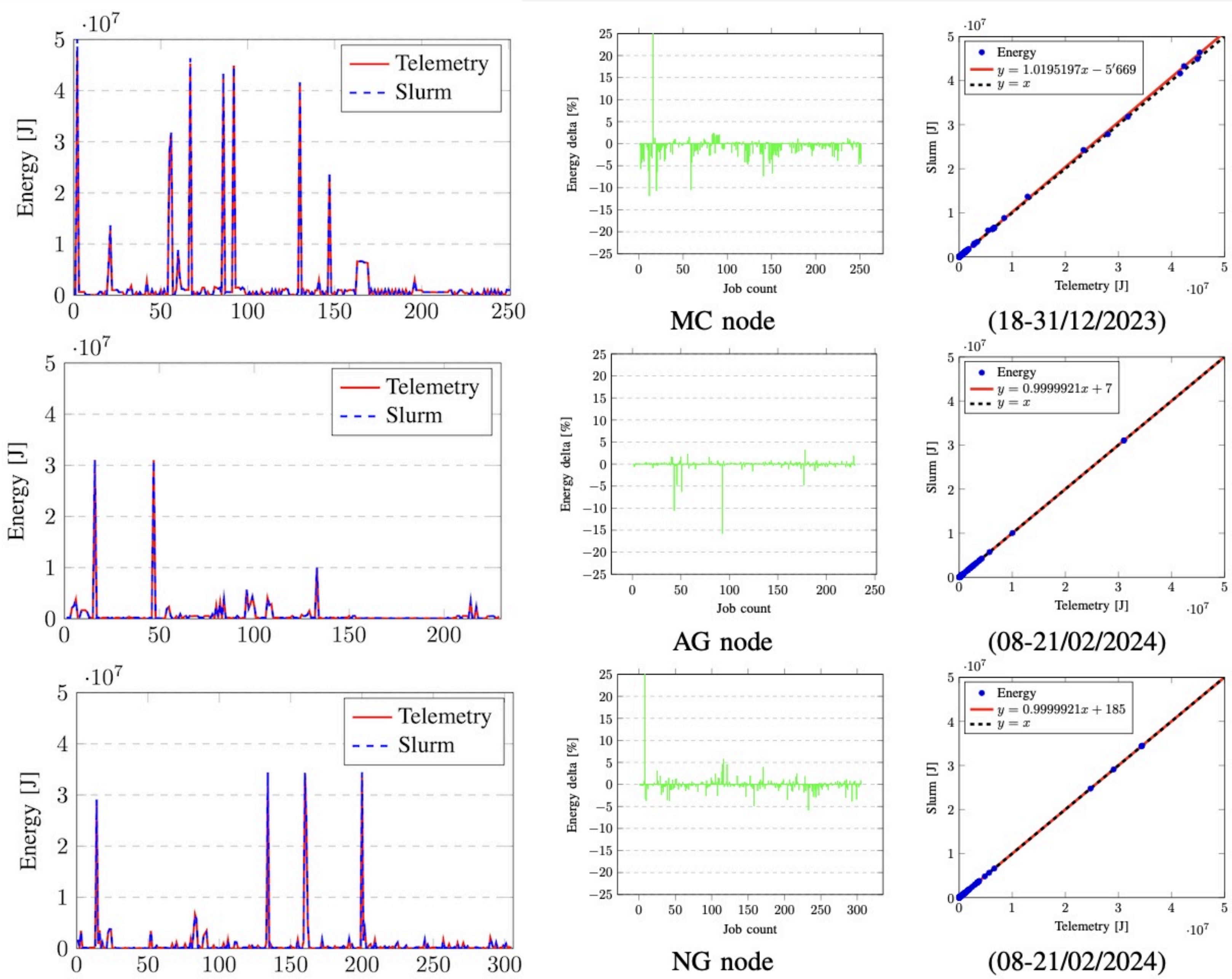


Validating Energy Data

- We validate data by comparing the energy data collected from slurm/pm_counters (sysfs) with the data collected from telemetry (redfish).

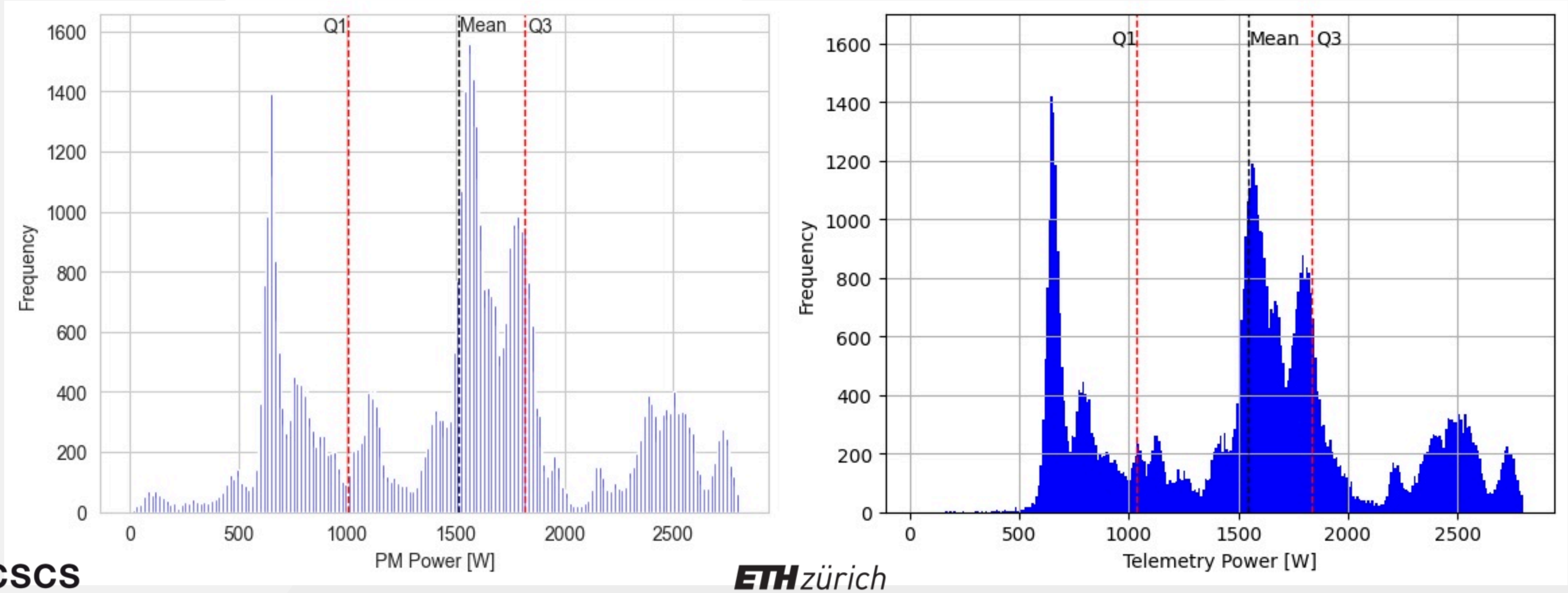
TABLE III: Alps Compute Node Specifications

Blade architecture	CPU(s) per node	GPU(s) per node
EX-425 Windom (MC)	2 AMD 64-core 7742	0
EX-325A Bard Peak (AG)	1 AMD 64-core 7A53	4 AMD MI200
EX-325N Grizzly Peak (NG)	1 AMD 64-core 7713	4 NVIDIA A100
EX Blanka Peak (GH)	1 ARM 288-core Grace	4 NVIDIA GH200



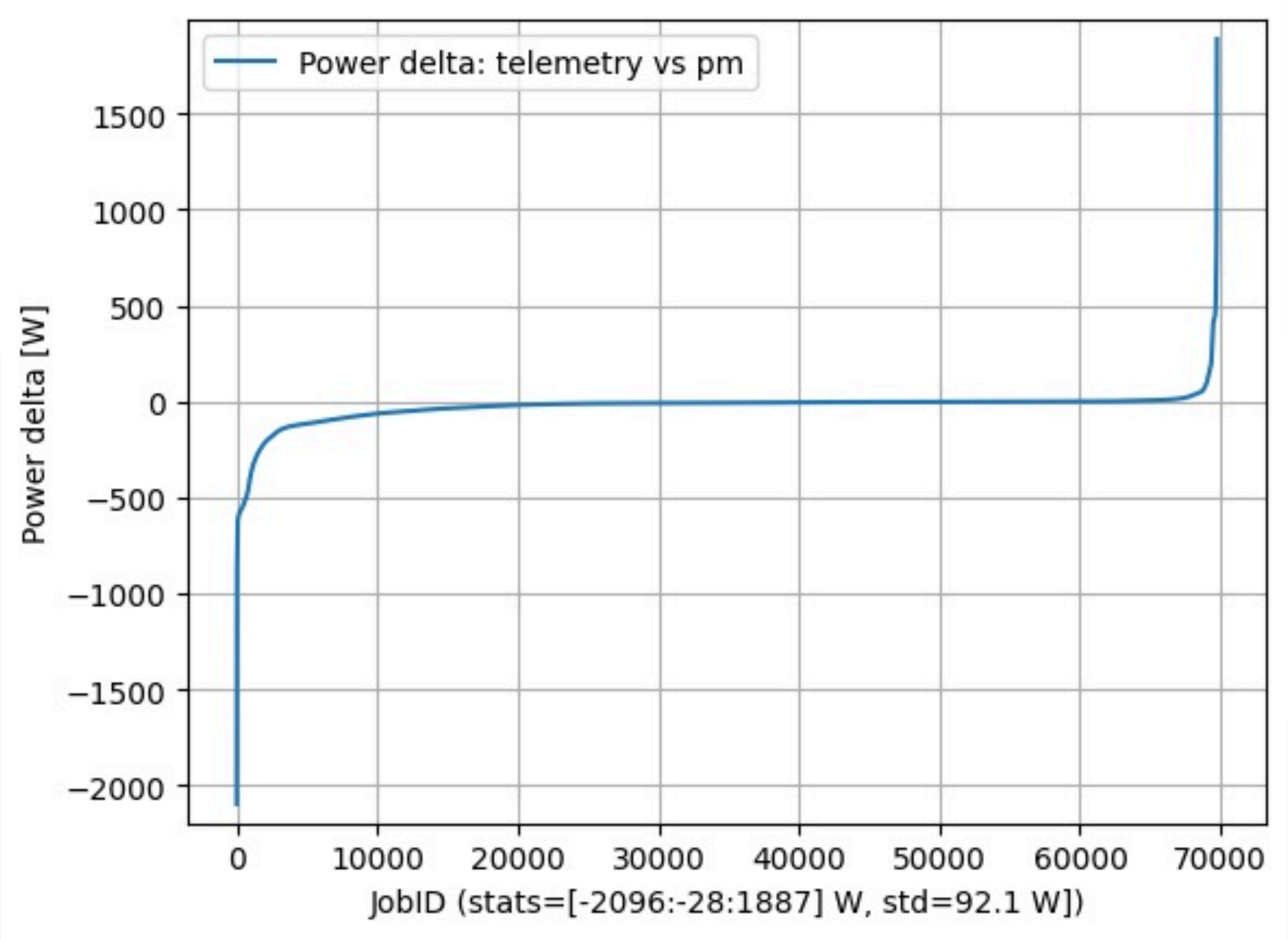
Validating Energy Data: Grace Hopper

- Cleaning 3 months of data (between Feb and May 2024) from outliers by removing jobs with:
  - more than 1 node (nnodes == 1) and short runtime (duration < 10sec),
  - null or unrealistically high energy (ConsumedEnergy == 0 or > 1e9 J),
  - unrealistically high power (Power > 2800 W),
- From **130,840** jobs to **83,845** jobs: a good mix of small, medium, and large power-intensive jobs / 64% of all recorded jobs

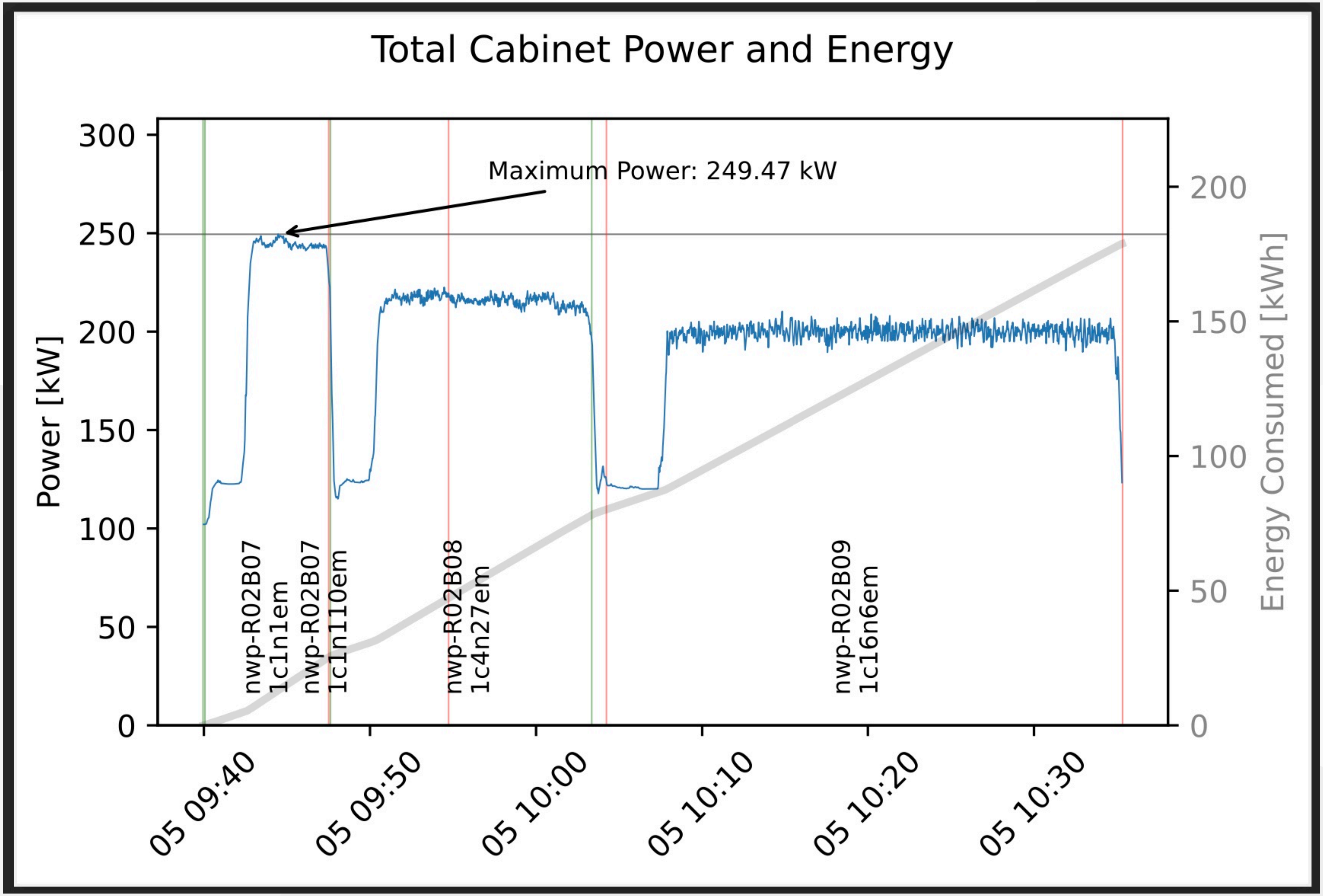


"Lies, damn lies and statistics"

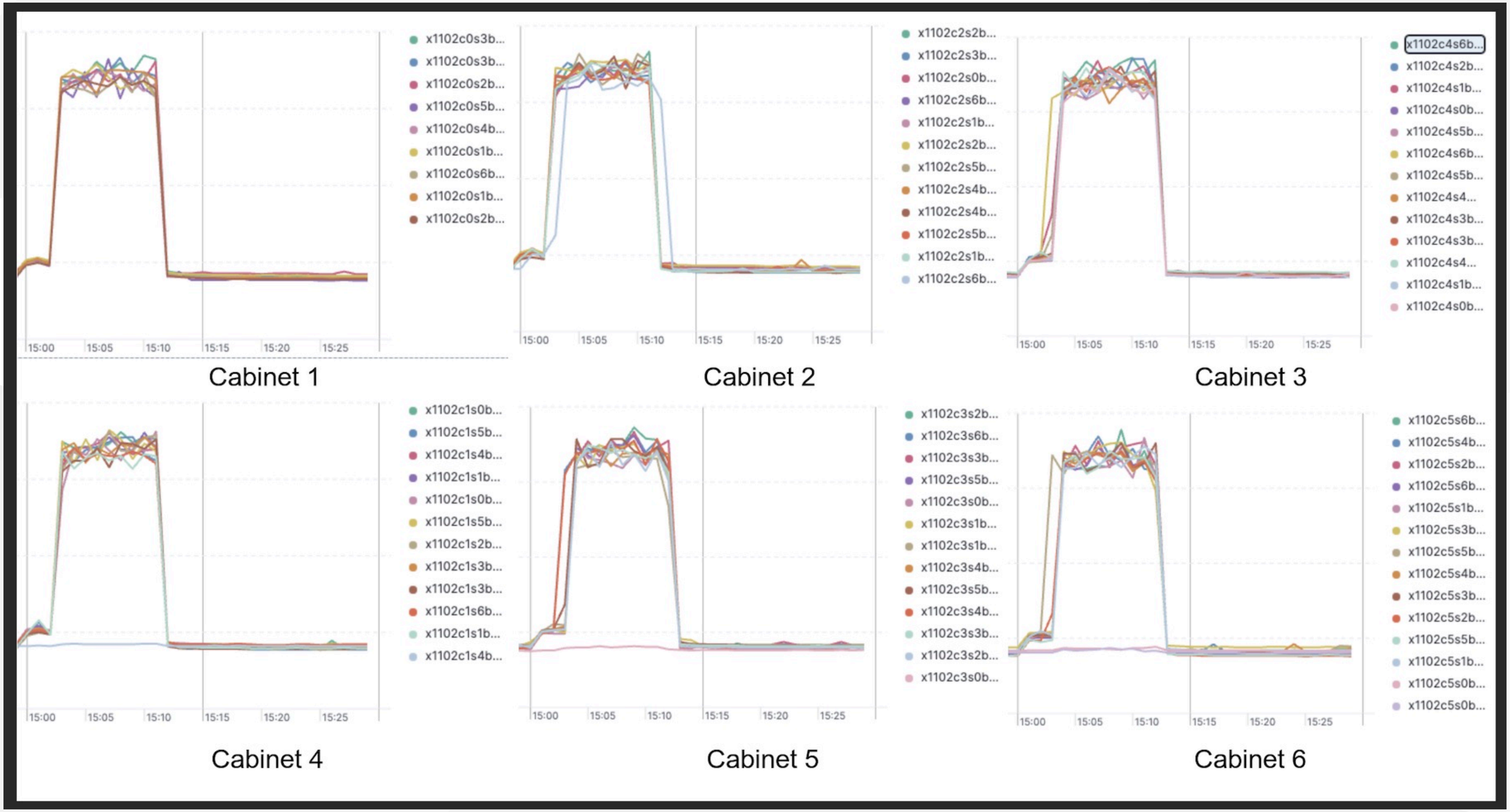
- Small number of discrepancies, where 1% of the jobs are showing an absolute delta > 500 W, these variations are under investigation.



GH Cabinet Power (112 compute nodes)



GH Row Power (6 cabinets)

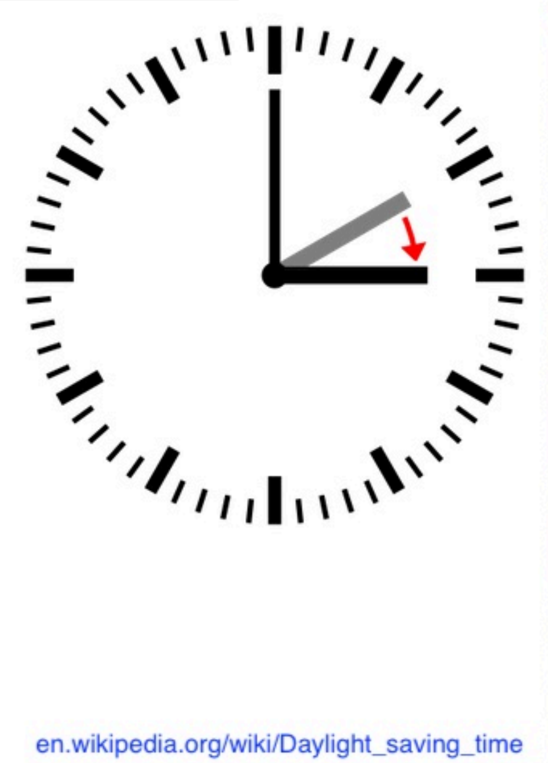
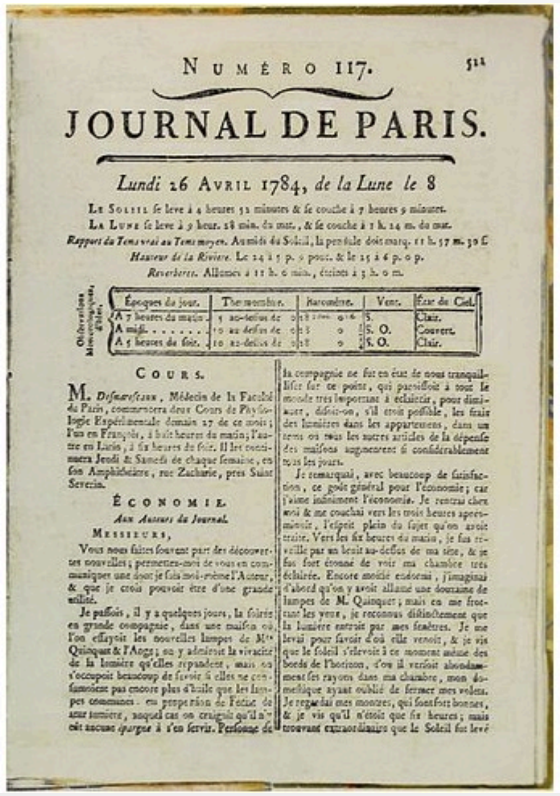


CSCS

ETH zürich

Central European Summer Time (CEST)

- Daylight saving time: advance clocks to make better use of the longer daylight available during summer
  - Proposed by Benjamin Franklin in April 1784,
  - Germany first country to implement it nation-wide in 1916,
  - From last Sunday in March to last Sunday in October (EU),
  - Interesting clock synchronization problem between facility/hpe/elastic tools.



NUMÉRO 117.

JOURNAL DE PARIS.

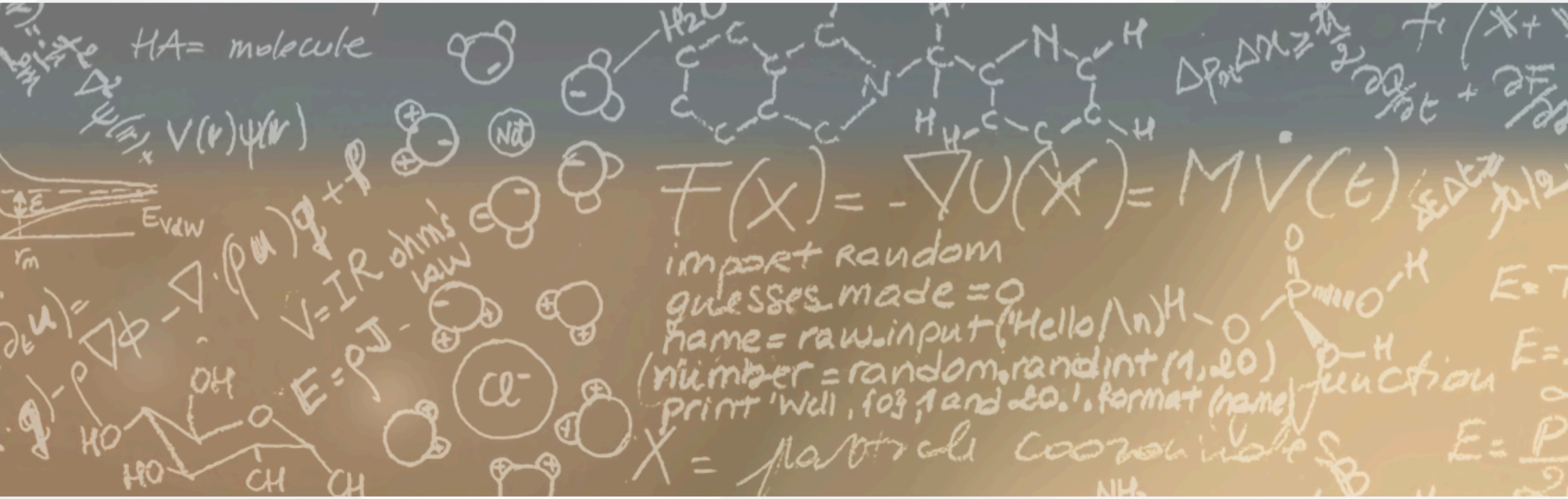
Lundi 26 Avril 1784, de la Lune le 8

en.wikipedia.org/wiki/Daylight_saving_time

Conclusion

- EMOI: CSCS Extensible Monitoring and Observability Infrastructure
  - Integration of CSM/SMA into EMOI,
  - Kafka-centric model with low overhead,
  - Git-ops approach is giving us flexibility to create/destroy clusters on demand.
  - Use Energy and Power data to encourage user to optimize their code.

The Data Warehouse and Data Intelligence (DWDI) team



Thank you