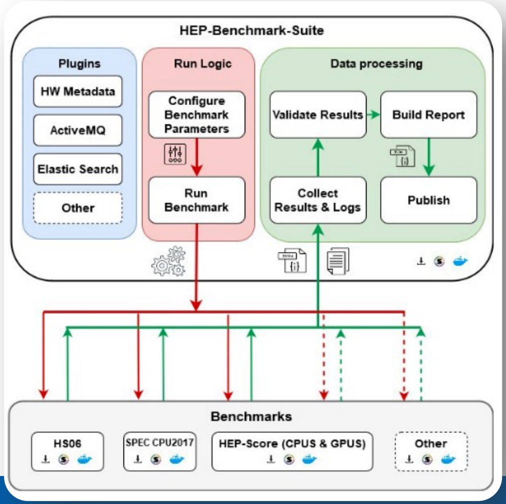
# Measuring Machine Metrics and Performance

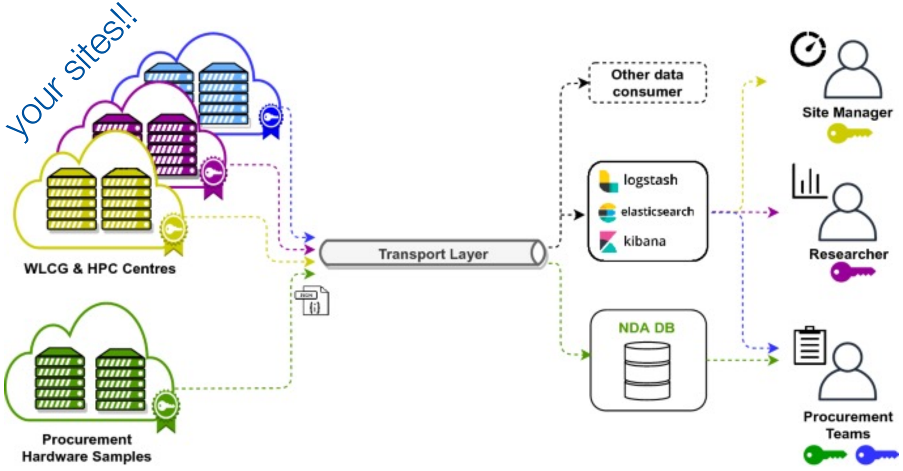
D. Giordano (CERN)
on behalf of
HEPiX Benchmarking WG

HEPiX Spring 2024
16/04/2024

# Recap from previous talk

- HEP Benchmark Suite (link)
  - Orchestrator of multiple benchmark (HEPScore, HS06, SPEC CPU2017)
  - Central collection of benchmark results. Reports have a modular JSON structure
    - Details about the running workloads
    - *Rich metadata information about the servers*



# Collect utilization metrics (I)

Example: Plugin Configuration

Suite expanded to collect timeseries utilization metrics alongside the running benchmark

- Flexible configuration approach
  – Command + regex
  – Sampling intervals

**Suite configuration**

```yaml
plugins:
  CommandExecutor:
    metrics:
      cpu-frequency:
        command: cpupower frequency-info -f
        regex: 'current CPU frequency: (?P<value>\d+).*'
        unit: kHz
        interval_mins: 1
      power-consumption:
        command: >
          sudo ipmitool sensor get 'PS1 Power In' ; sudo ipmitool sensor get
          'PS2 Power In'
        regex: 'Sensor Reading\s+:\s*(?P<value>\d+).*'
        unit: W
        interval_mins: 1
      load:
        command: uptime
        regex: 'load average: (?P<value>\d+.\d+),'
        unit: ''
        interval_mins: 1
      used-memory:
        command: free -m
        regex: 'Mem: *(\d+) *(?P<value>\d+).*'
```

# Collect utilization metrics (II)

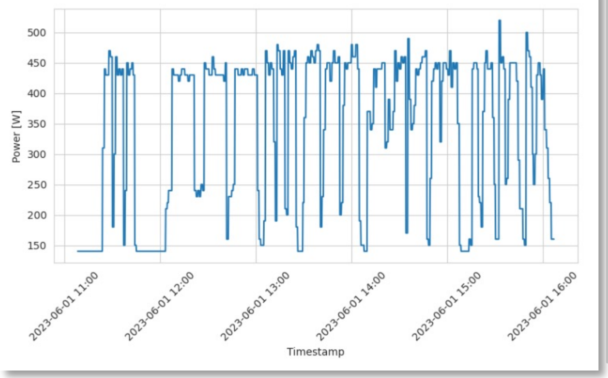
Plugins' Report

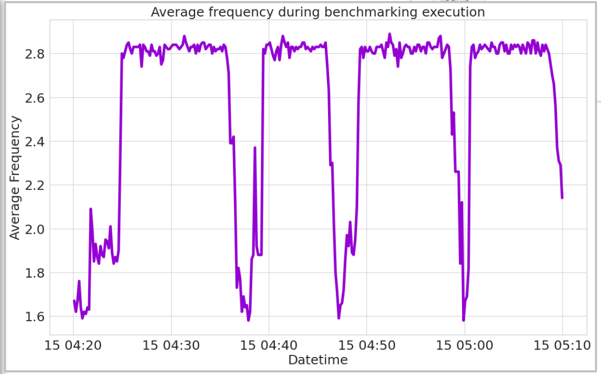

```
"plugins": {
    "CommandExecutor": {
        "hepscore": {↔},
        "pre": {
            "load": {
                "start_time": "2023-09-03T11:19:08.772095Z",
                "config": {
                    "command": "uptime",
                    "interval_mins": 1,
                    "aggregation": "sum",
                    "regex": "load average: (?P<value>\\d+.\\d+),",
                    "unit": ""
                },
                "values": [11.05,10.47,10.49,10.22,10.45,10.54],
                "end_time": "2023-09-03T11:24:08.774851Z",
                "statistics": {
                    "min": 10.22,
                    "mean": 10.536,
                    "max": 11.05
                }
            },
            "status": "success",
            "used-memory": {↔},
            "used-swap-memory": {↔}
        },
        "post": {↔}
```


## Time Series report

- Individual measurements for deep analysis
- Aggregated statistics for prompt visualization

Retrieved timeseries from stored data



# HEPScore23 + Usage metrics Examples:

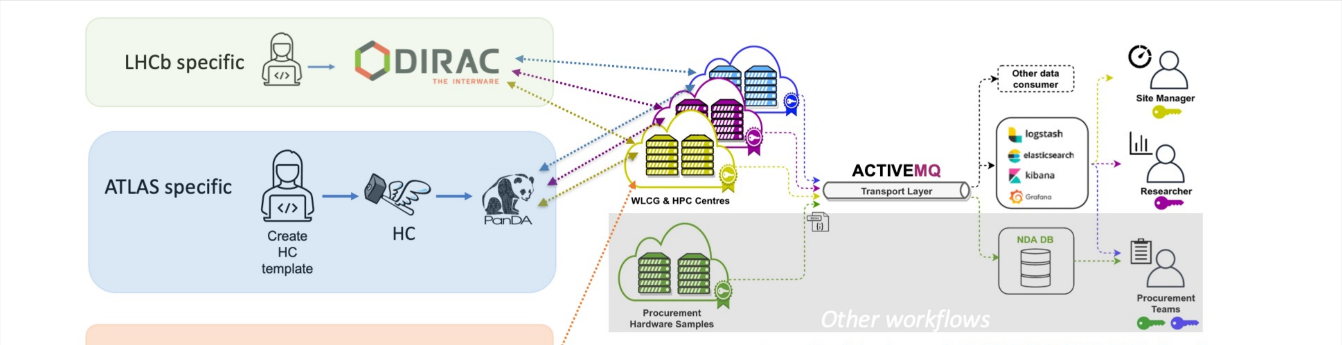
- Probe grid compute performance
- Power consumption and environmental impact

# Probe grid compute performance

- Use HEPScore23 to probe the performance of WLCG **job slots**
  - Access workload images from cvmfs
- Use the HEP Benchmark suite to correlate performance with server utilization metrics
  - Load, memory usage, power consumption, etc
- Inject the probe as a normal job via the workload management system of the experiment (PanDA, Dirac, HammerCloud, etc)
- Work started last year in collaboration with ATLAS
  - Extendable to other VOs (already exercised with LHCb)
  - Status presented at the January GDB and at ACAT

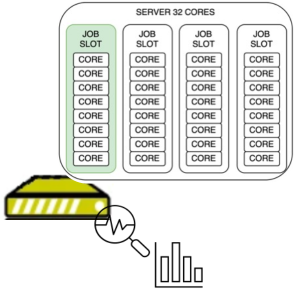
# In job slots as in bare-metal nodes

- The benchmarking process is injected in the site job slot via standard job submission systems
  - Probe multi-core job slots (8/4/1 cores)
- Same data flow used to collect HS23 data from bare-metal nodes
  - AMQ → logstash → OpenSearch & HDFS (monitoring and analysis)
- Successfully implemented and deployed the pipeline for:
  - ATLAS: Automated submission via HammerCloud
  - LHCb: Manual submission to DIRAC




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

Benchmark execution as payload of a job slot




*Collection of the server utilization metrics*

# Data collected

## ATLAS

data from: 07/07/23 – 08/04/24

- Automated job submission every 3 hours on each panda resource
  - 111 Panda Resources
  - 163 CPU Models
  - 29162 unique hosts
- Over 100k successful jobs
- Each job: 8 core slot
- Median of job's walltime: 83 minutes
  - HEPScore23 configuration with 1 repetition
  - 0.06% of total walltime_x_core

## LHCb

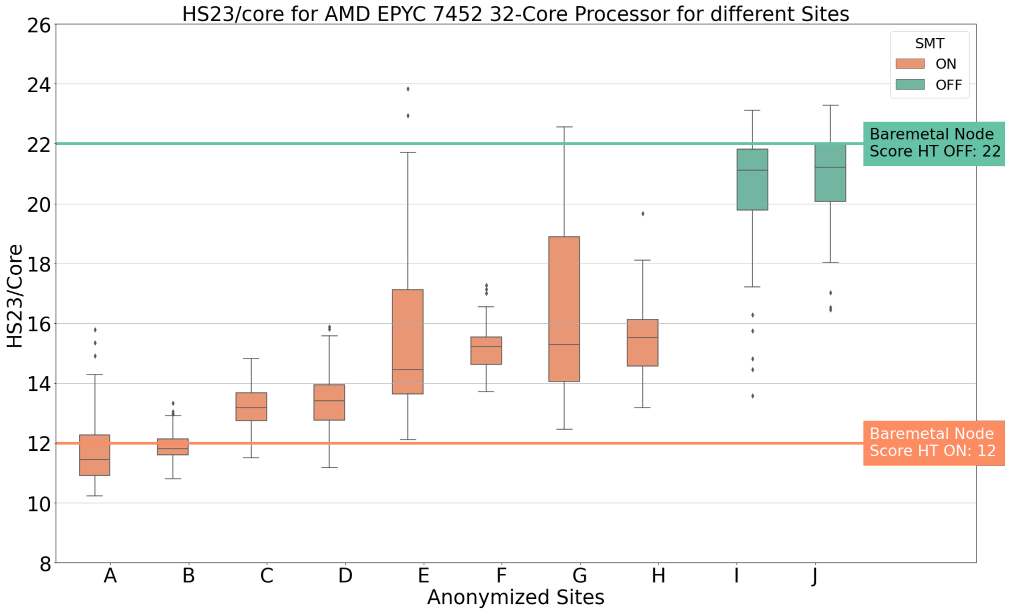
data from: 01/08/23 – 01/11/23

- Manual job submission
  - 48 Sites
  - 110 CPU Models
  - 1650 unique hosts
- 2.1k jobs finished
- Each job: 1 or 4 core slot (most 1core)
- Median of job's walltime: 43minutes
  - lhcb-sim-run3-ma-bmk with 3 repetitions

# First evidence

Servers with the same CPU model can perform very differently from grid site to grid site

Real performance per core delivered to the job slot

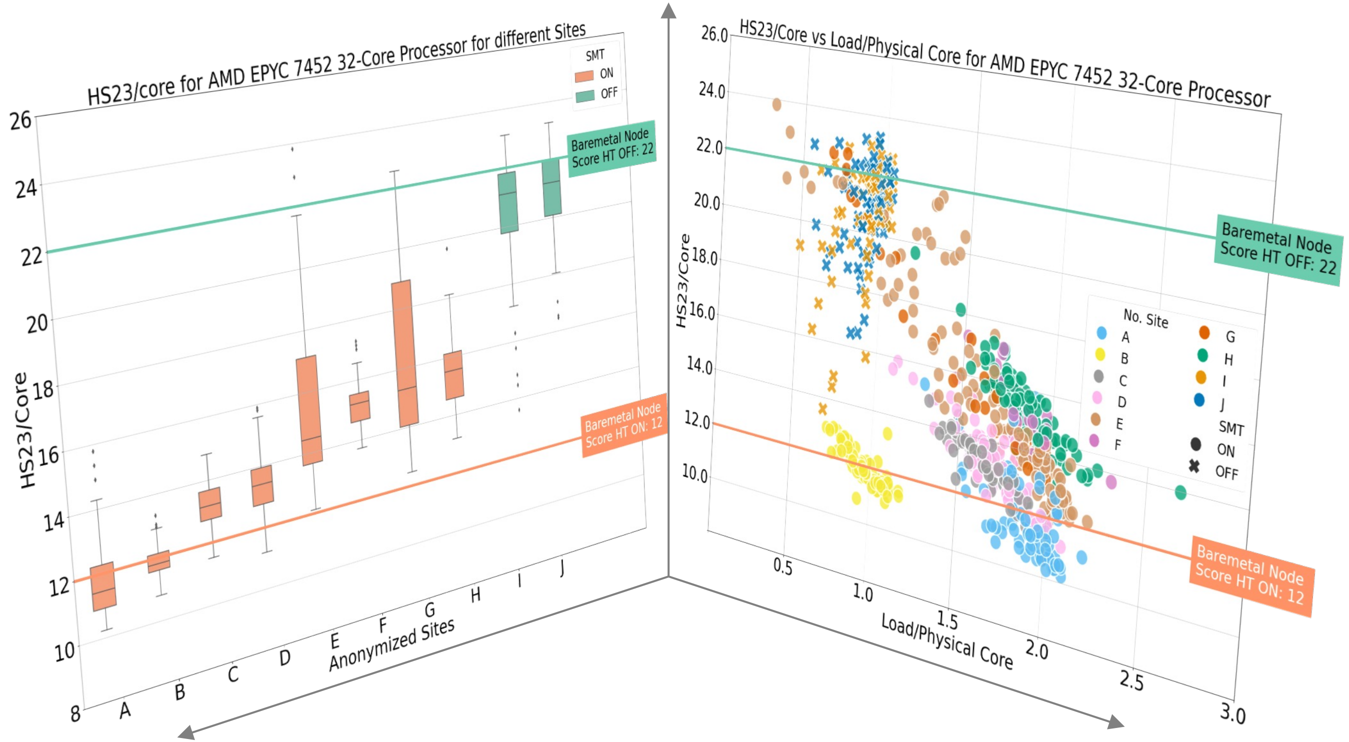



*Natalia Szczepanek*
*Atlas SW & Computing week Jun 2023*

Main cause can be explained considering the server status at the measurement time

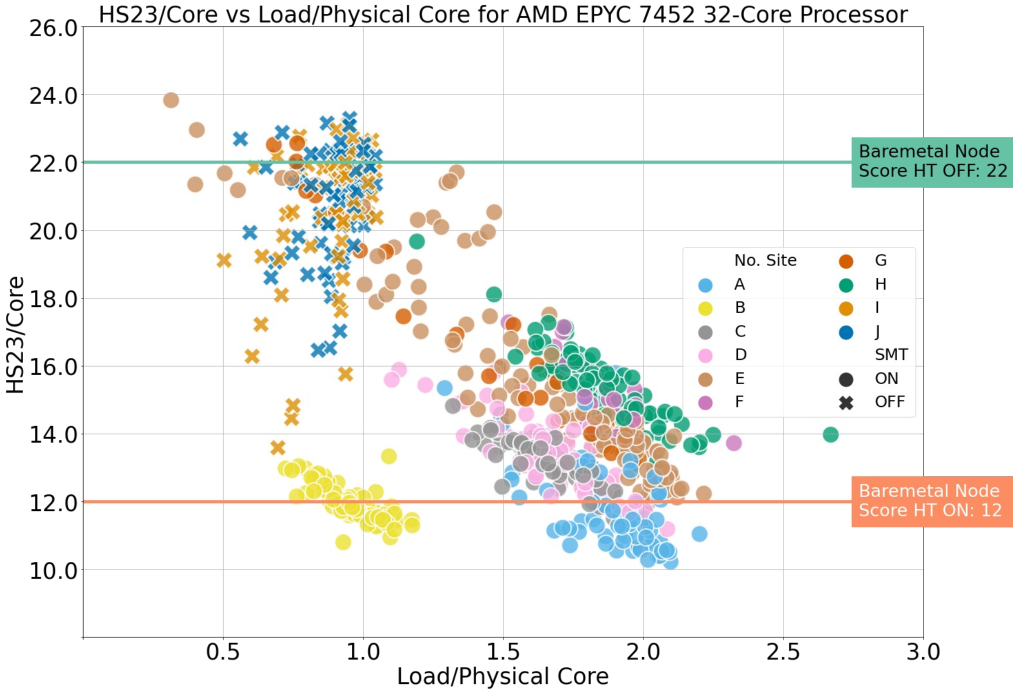
# Expanding the feature space

## CPU load justifies the main trend and helps spotting anomalies

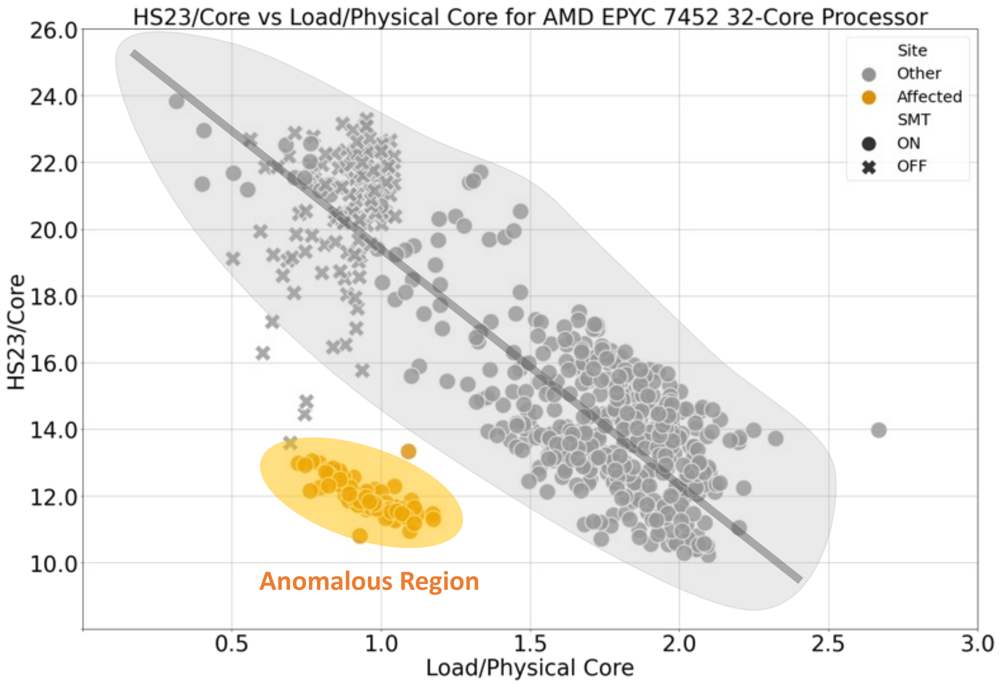


# Expanding the feature space

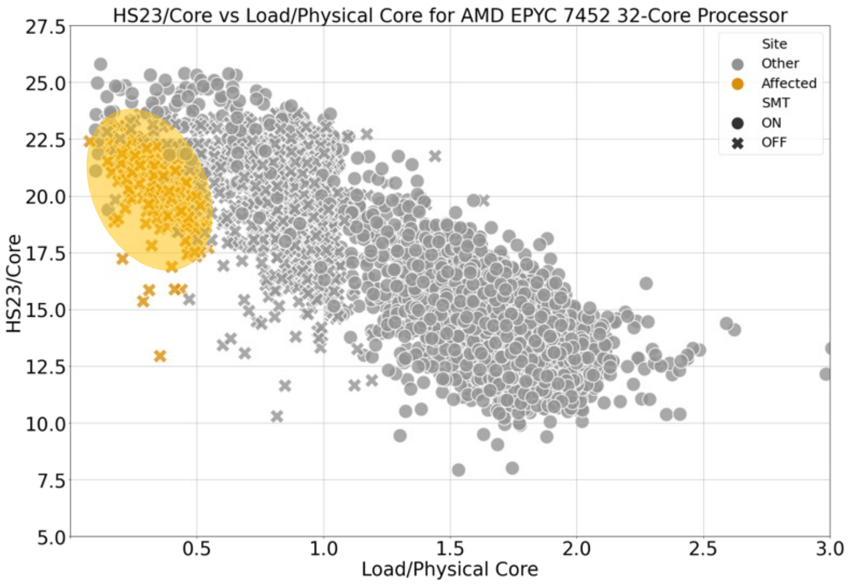
## CPU load justifies the main trend and helps spotting anomalies



# Fixing misconfiguration issues




*After configuration fix in the nodes (done by sysadmins), the performance of the affected site increased by 66%*

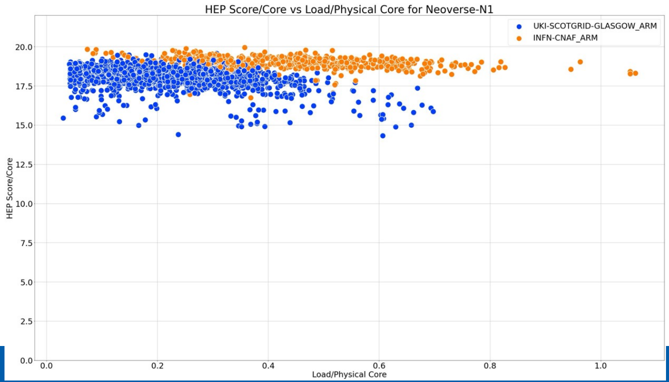


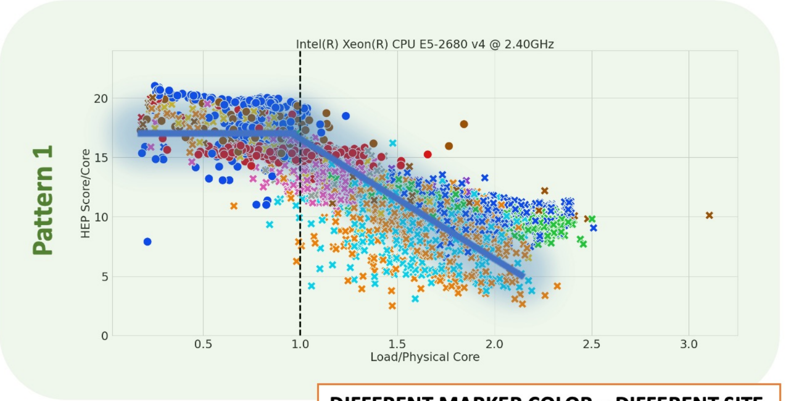
# Build a data model

Enriched data give the opportunity to build better models

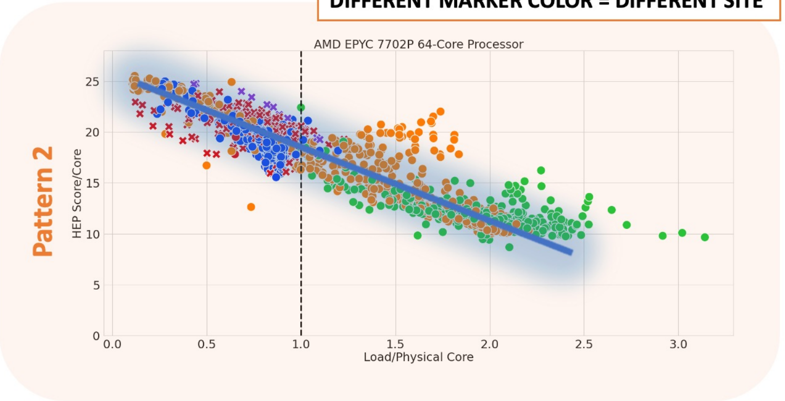
- Two main patterns on x86
  - Not strictly correlated to Intel Vs AMD
  - **Pattern 1**:consistent with the thread scan results
  - **Pattern 2**: additional performance boost.
    Still not clear the reason, effect of a second feature, or modernization of the CPU

- Different patter for ARM, no SMT



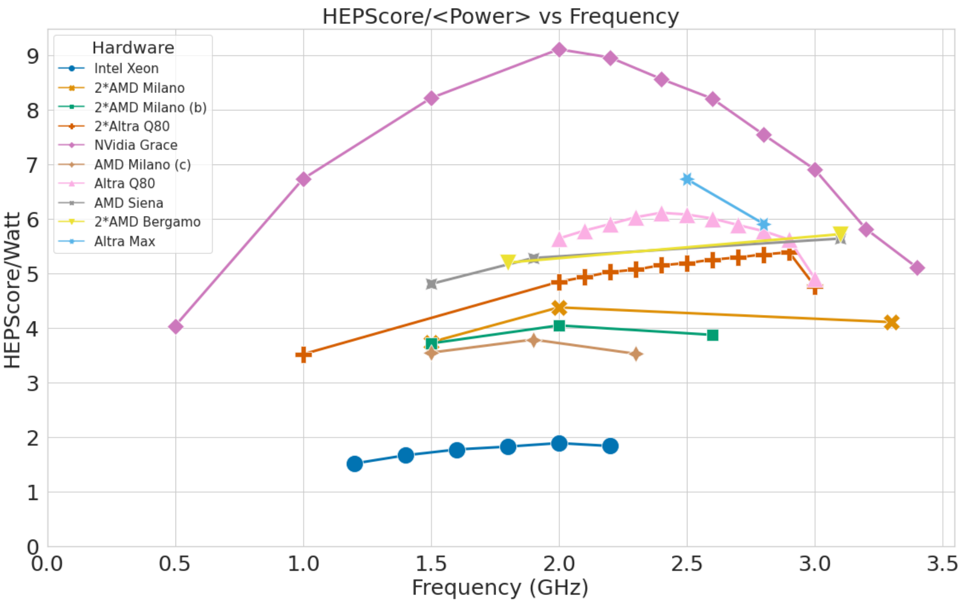




**DIFFERENT MARKER COLOR = DIFFERENT SITE**



# Another use case: Power Consumption Study

- Power measurement together with frequency and HEPScore
  - Requires access to the entire server to read power consumption
  - Not possible via grid job submission
- Preliminary findings
  - HEPScore/Watt Vs frequency shows different characteristic curves depending on the CPU model and architecture
- Work in progress: would benefit from other contributors




*E. Simili ScotGrid Glasgow*

# Summary

- HEP Benchmark Suite is a powerful tool not only for benchmarking the entire server
  - Enhanced to include server utilization metrics
  - Multiple opportunities for studies
    - Model HS23 vs (Load, other metrics) as a calibration tool
    - Anomaly detection: monitor site performance and fix misconfigurations
    - Performance/Watt (power consumption) and GPU utilization studies just started
    - Others

CERN