

ABOUT THE PROJECT

- **Goal:** Provide cluster operators and users with an automated rule-based detection system for known pathological HPC jobs (jobs with previously encountered inefficiencies).
- **Duration:** January 2023 - June 2024 (extended)
- **Funding:** German National High Performance Computing

APPROACH

Data Sources



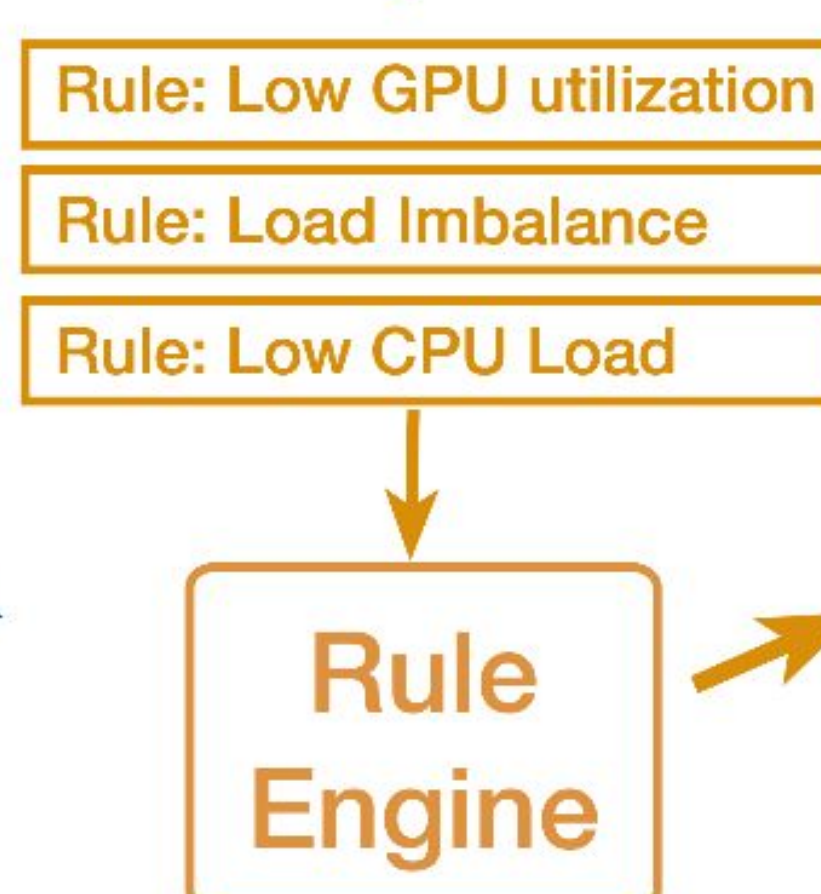
Storage



Job Archive

- Hardware counter series
- Job Information
- Runtime Data
- Binary Identification

Automated Analysis



Outbound Action



HPC user

- Cluster Cockpit generates a job-archive for each job; alternative sources experimented with, but not ready to use
- Rule-Engine evaluates each anti-pattern and triggers appropriate for each job archive
 - Anti-pattern cluster relative hardware and software metrics in conjunction with binary identification methods indicating inefficient behaviour
- Detected anti-patterns trigger action templates (E-Mail), notifying users or staff depending on the setup

CURRENT PERFORMANCE ANTI-PATTERNS

- Low CPU load
- Load-Imbalance Pattern
- Job Payload Overhead Pattern
- Excessive CPU load
- Problematic memory usage
- Uncoordinated multi-process GPU-usage
- Low GPU utilization
- Crypto Miner
- Unnecessary job distribution
- Unnecessary PFS read

RULE EXAMPLE

```
{
  "name": "Low CPU load",
  "tag": "lowload",
  "parameters": ["threshold_factor"],
  "metrics": ["cpu_load", "job"],
  "terms": [
    { "load_mean": "cpu_load.mean('all')"},
    { "load_threshold": "job.numHwthreads * threshold_factor"},
    { "lowload_nodes": "load_mean < load_threshold"},
    { "lowload": "lowload_nodes.any('all')"},
    { "load_perc": "1.0 - (load_mean / load_threshold)"}],
  "output": "lowload",
  "output_scalar": "load_perc",
  "template": "Job ({{job.jobId}}) was detected as the mean cpu load {{load_mean}} falls below {{load_threshold}}."
}
```

RESULTS & FIRST EXPERIENCE

Deployment:

- The "PathoJobs" system is deployed by partners on production HPC systems, running with the Cluster Cockpit on dedicated VMs.
- Each instance processes job-archives from hundred-thousands to millions of jobs per month
- Current performance anti-patterns are evaluated for all HPC jobs

First Experience & Insights:

- Rule evaluation does not exhibit noticeable extra-load on the Cluster Cockpit hosts
- Number of jobs with detected inefficiencies exceeds acceptable rates
- System is able to detect explicitly malformed test-jobs

Future work / Next steps:

- Thresholds for automated alerts to mitigate high alert rates
- Current system is limited to apply rules to a single job-archive: extension can evaluate across multiple job-archives (aka jobs)
- Injection of detected inefficiencies to job-archives for display in Cluster Cockpit

Location	Jobs / month	Jobs with issues	Rules with highest hit-rate	Analysis cost
TU Darmstadt	> 62.9 k	598 / 1000	Low-CPU load Load-imbalance	230 ms / job
Paderborn	> 260k	180 / 1000	Low-CPU load CPU oversubscription	560 ms / job

* www.clustercockpit.org