Job specific performance monitoring with ClusterCockpit

DKRZ TechTalk

Pay Giesselmann

Motivation:

Performance increases Throughput

Performant means Energy Efficient
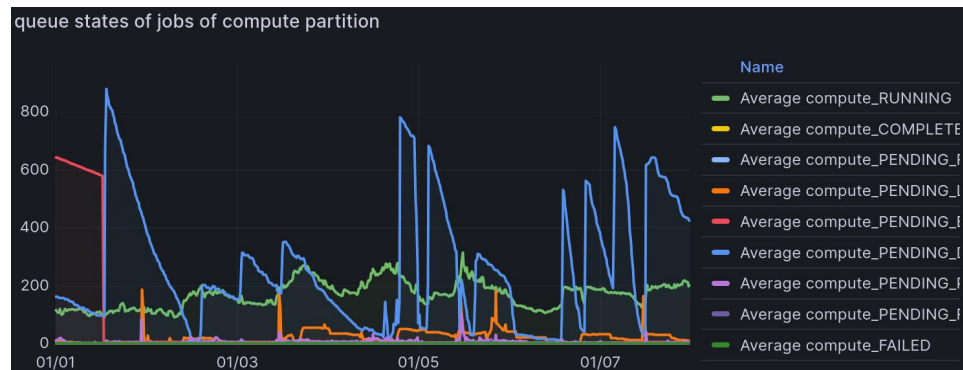
https://clustercockpit.dkrz.de

Agenda

- Guided Tour

- Performance Metrics

- Use Cases

Available Monitoring

- squeue -u <uid>

- scontrol show job <jobid>

- sacct -u <uid>

- log files ...

https://monitoring.dkrz.de

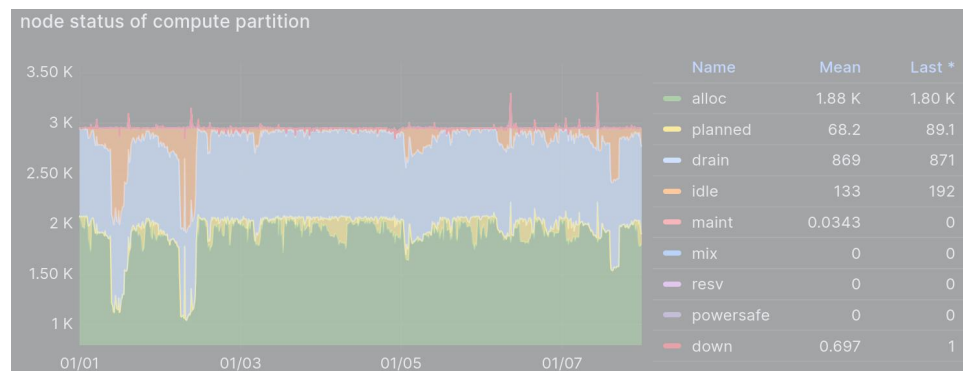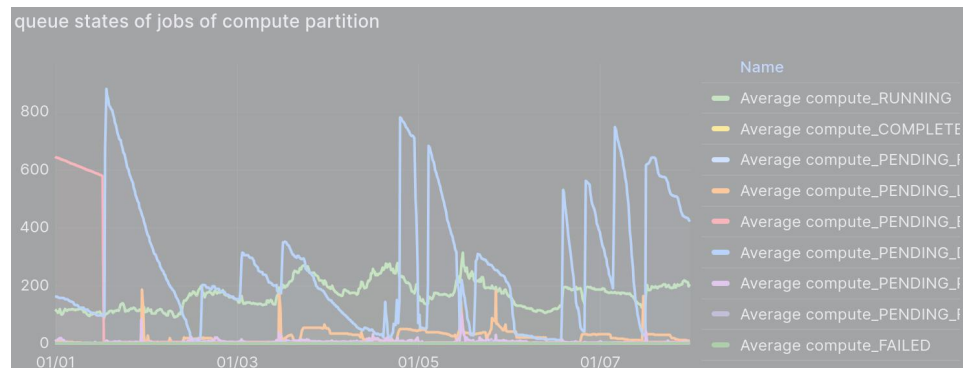# Available Monitoring

- squeue -u <uid>

- scontrol show job <jobid>

- sacct -u <uid>

- log files ...

- But what happened to my jobs during runtime?
  - Job specific?
  - Performance?
  - Resolved in time?

https://monitoring.dkrz.de

queue states of jobs of compute partition

Name
- Average compute_RUNNING
- Average compute_COMPLETE
- Average compute_PENDING_I
- Average compute_PENDING_I
- Average compute_PENDING_I
- Average compute_PENDING_I
- Average compute_PENDING_I
- Average compute_PENDING_I
- Average compute_FAILED

node status of compute partition

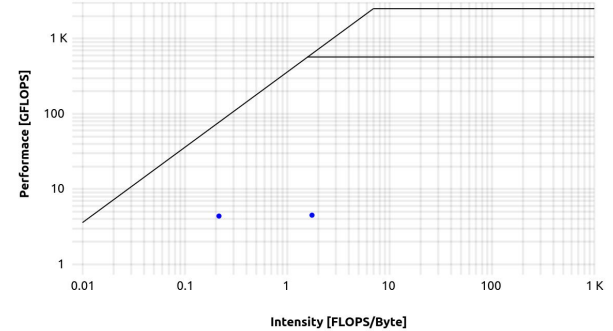| Name	Mean	Last *
alloc	1.88 K	1.80 K
planned	68.2	89.1
drain	869	871
idle	133	192
maint	0.0343	0
mix	0	0
resv	0	0
powersafe	0	0
down	0.697	1

Guided Tour

Guided Tour - My Jobs

Guided Tour - Job Details
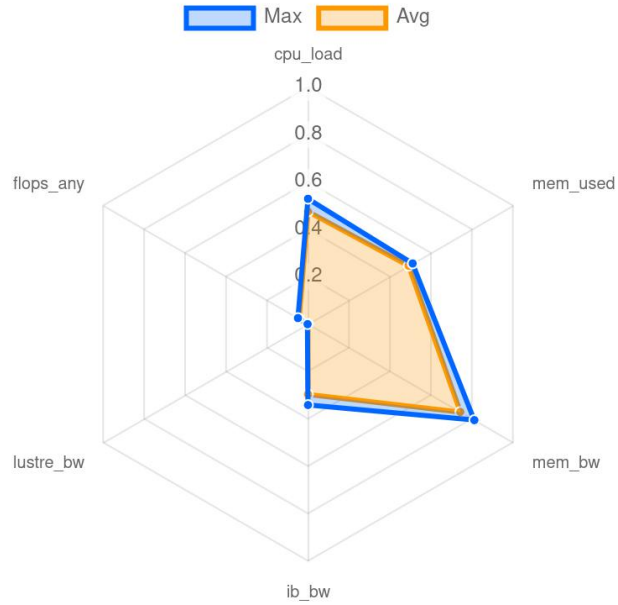
Footprint
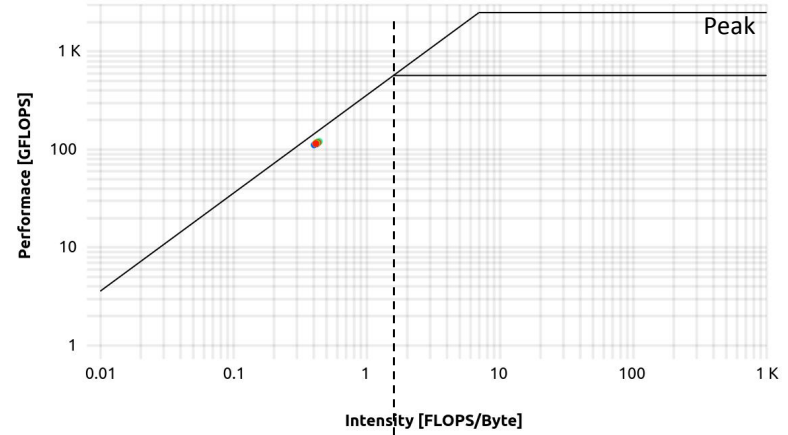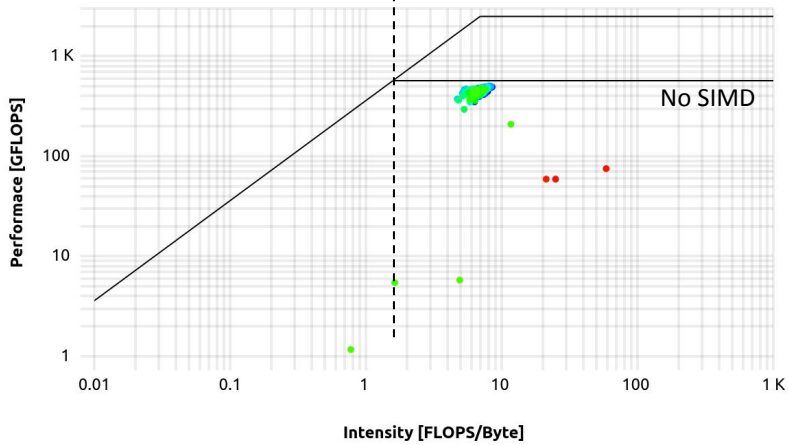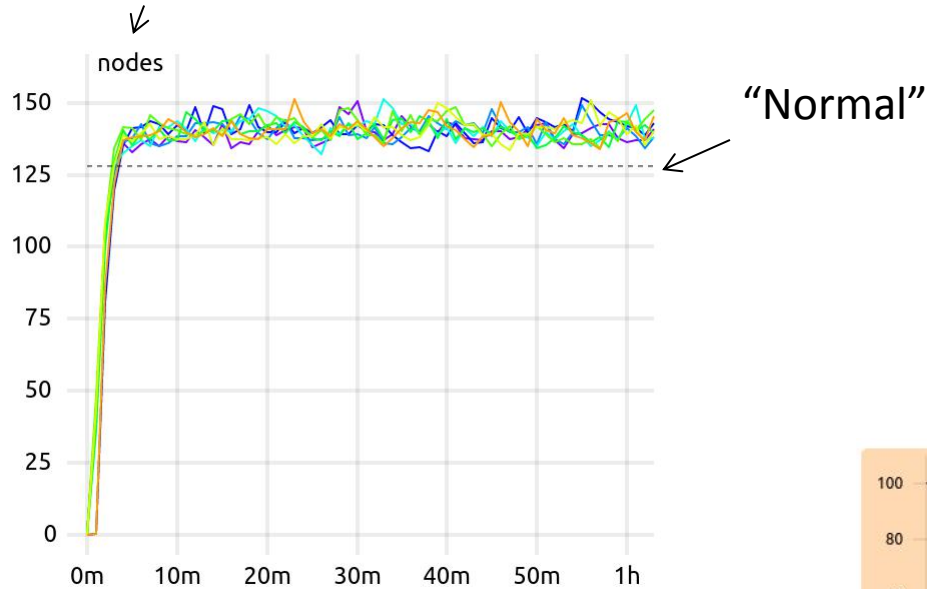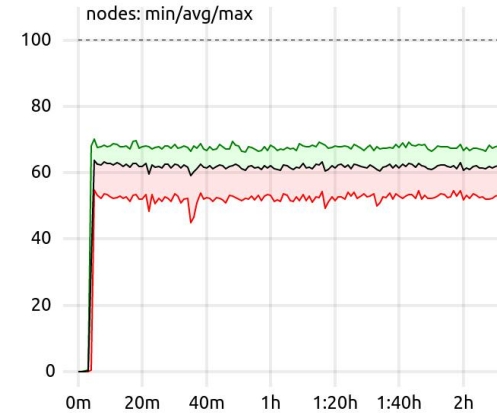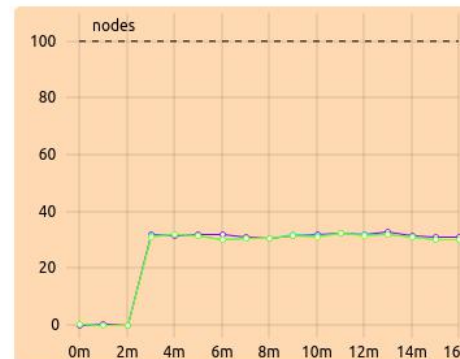
Memory bound

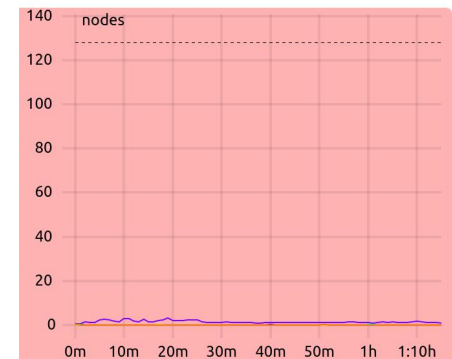Compute bound

Performance Metrics

Metric Scope

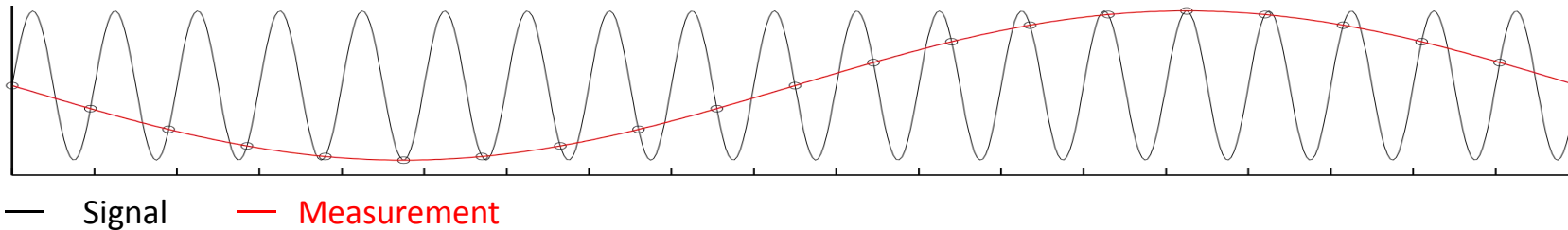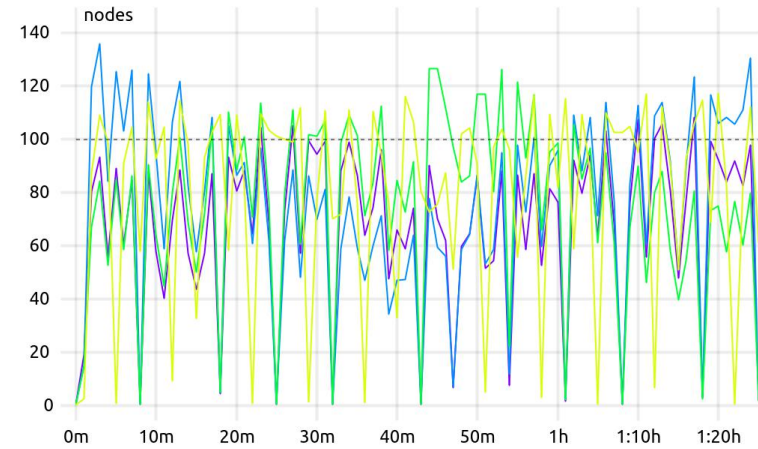"Normal"

Large Jobs > 15 nodes

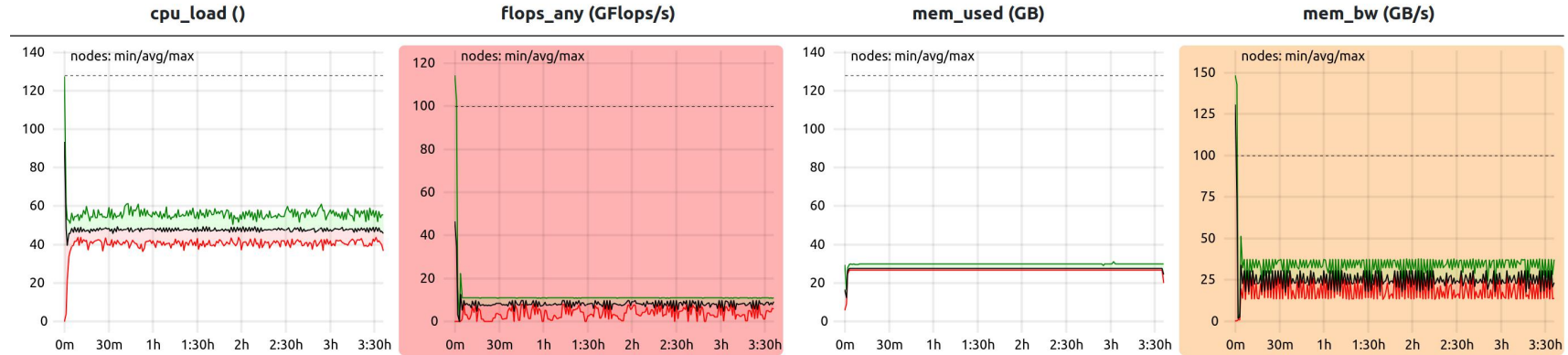FLOPS (Caution)

CPU Load (Alert)

Sampling Artefacts and Aliasing

- Fixed one minute sampling
- Async job scheduling

Signal — Measurement
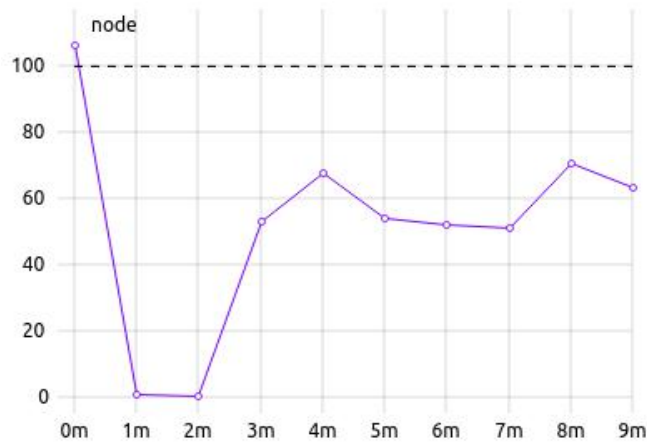
What is it good for?

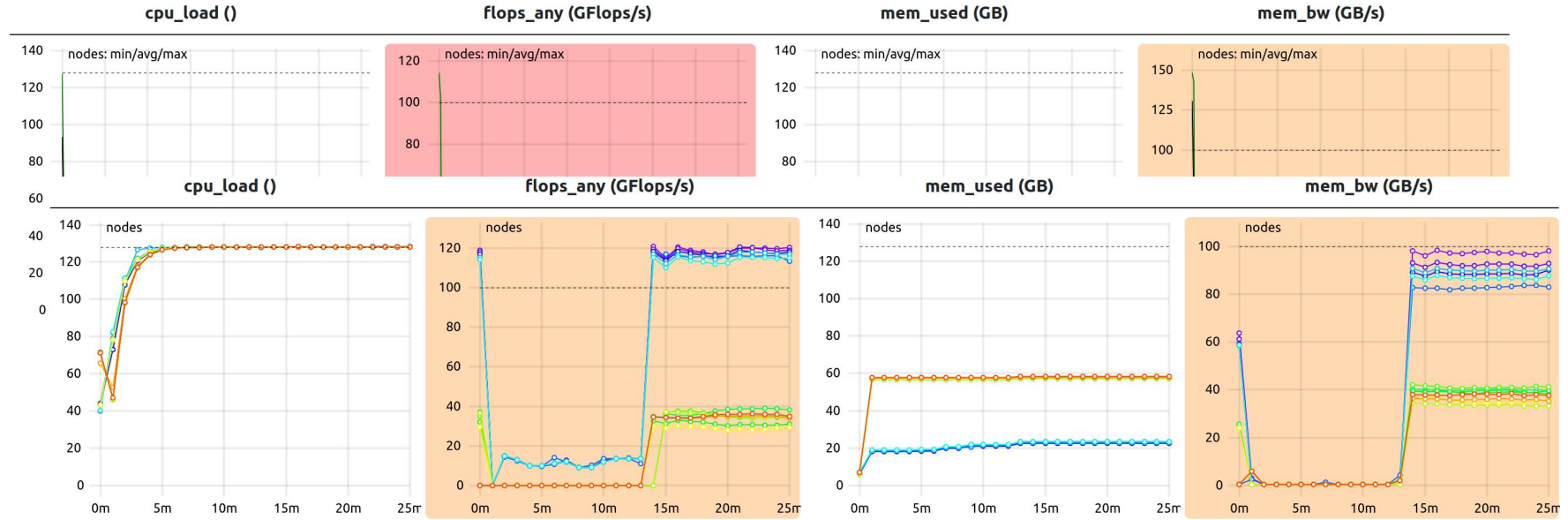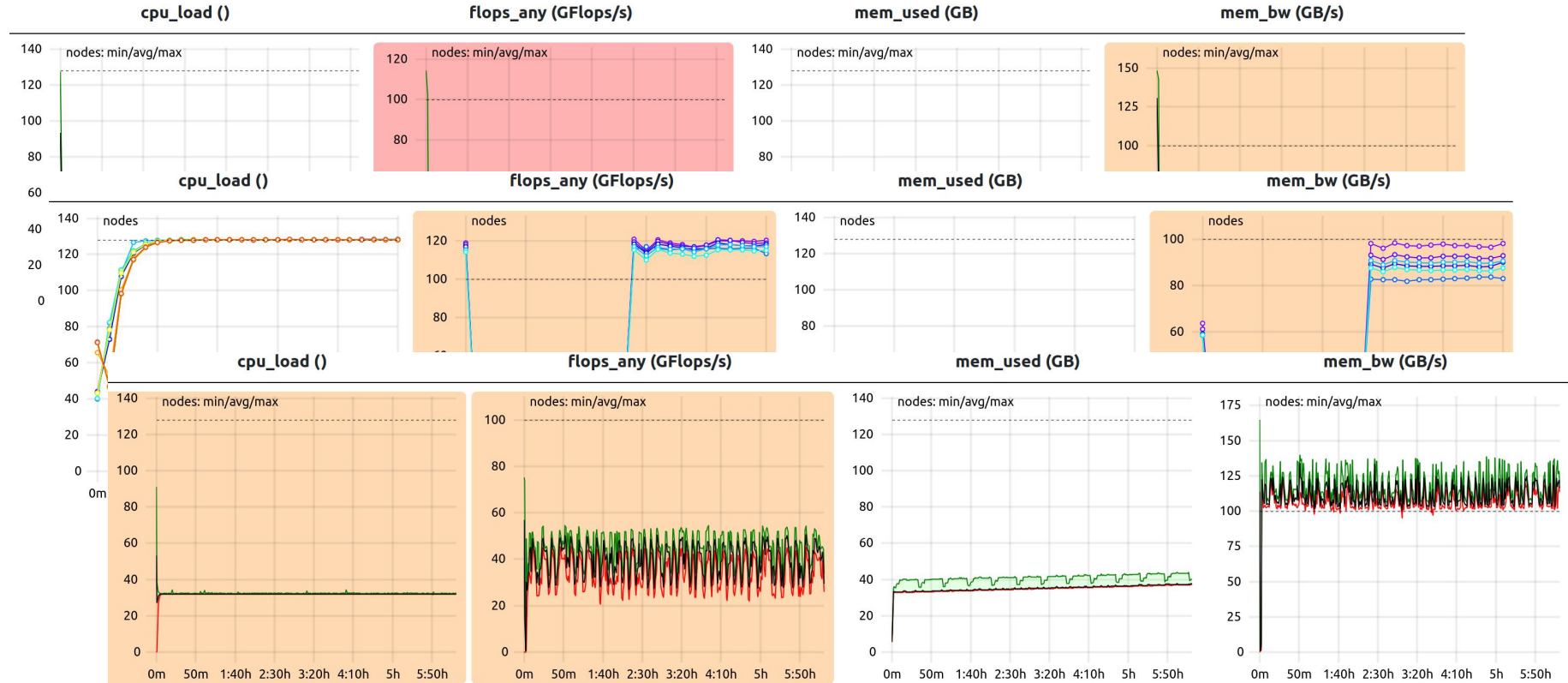# Intuitive Overview

101 nodes

101 nodes

Intuitive Overview

101 nodes

80 nodes, OMP_NUM_THREADS=4, missing libiomp5.so

Jobs:

> **5min** with

State **completed**

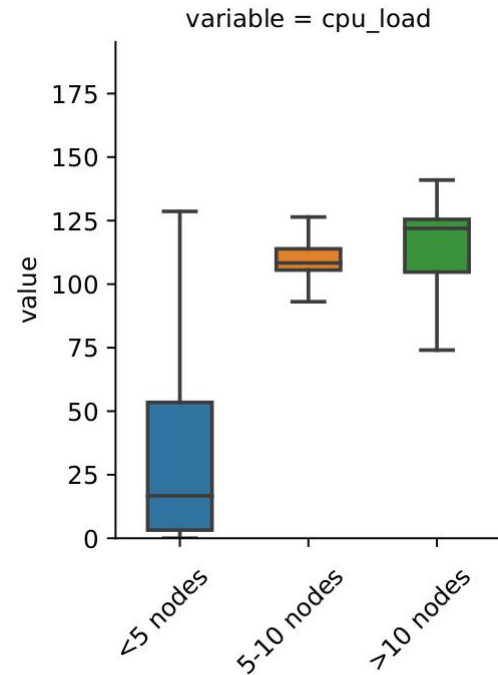From 01.01 to 07.01

Levante **compute** partition

Plots:

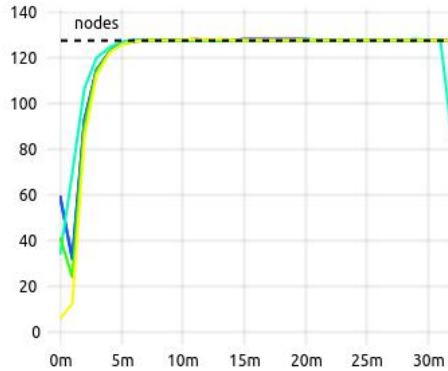Distribution of job averages

variable = cpu_load

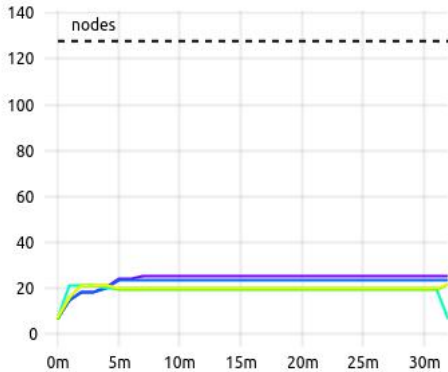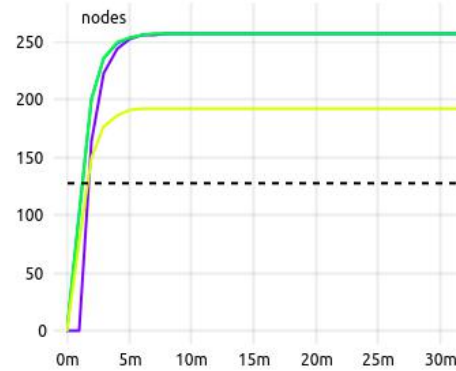| Category	Jobs	Node Hours
<5 nodes	11.329	19.477
5-10 nodes	5.617	35.087
>10 nodes	1.948	225.773

Performance Goals - On Point

128 Cores

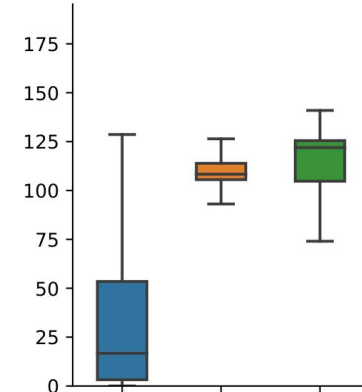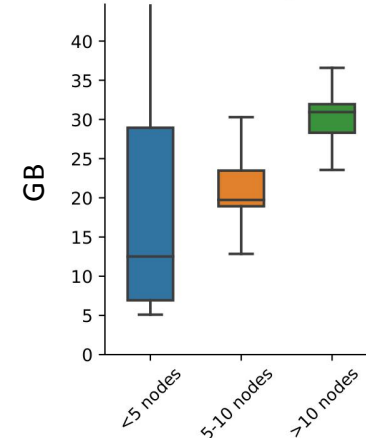**256 HW-Threads**

Levante:
256 GB   (2670 nodes)
512 GB   (294 nodes)
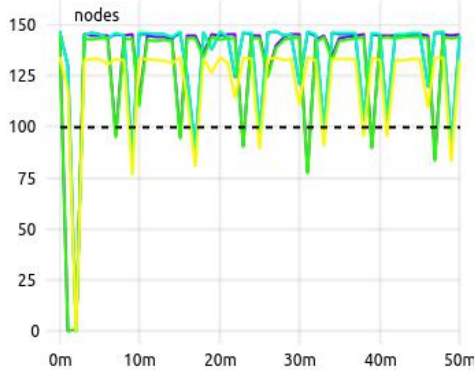1024 GB   (18 nodes)

variable = cpu_load

variable = mem_used
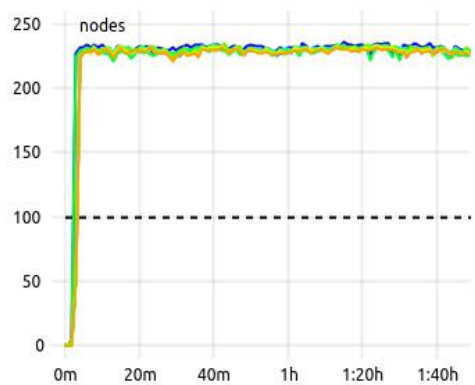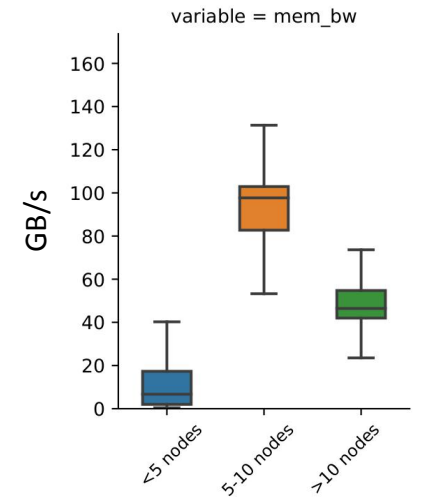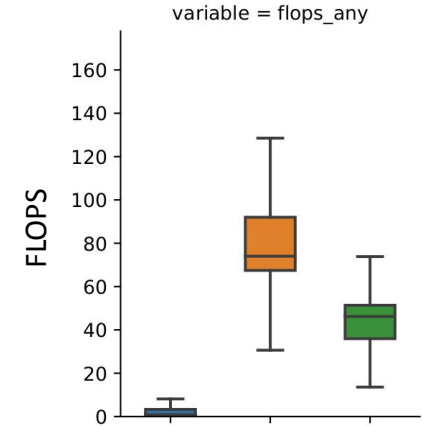
Performance Goals - Let It Burn

NEMO 4.2 @5 nodes

ICON @8 nodes
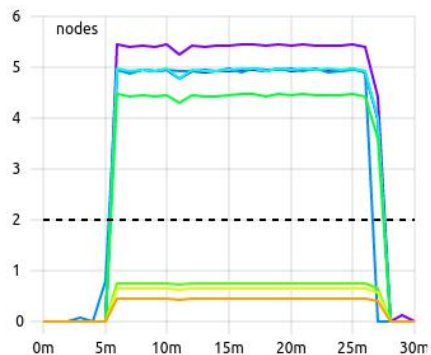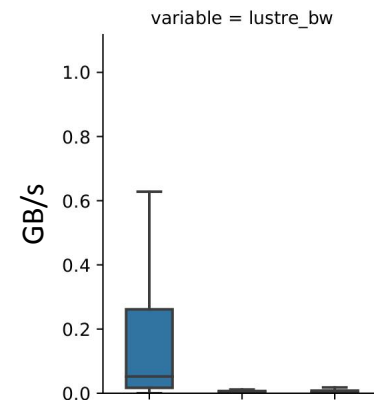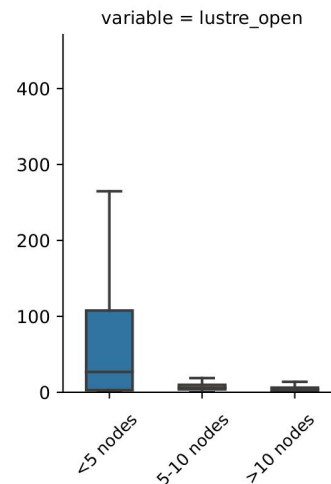

LIKWID Bench:
570 GFLOPS (Scalar)
2500 GFLOPS (SIMD)

STREAM:
360 GB/s

variable = flops_any

variable = mem_bw

Performance Goals - As Much As Needed

lustre_open

lustre_bw
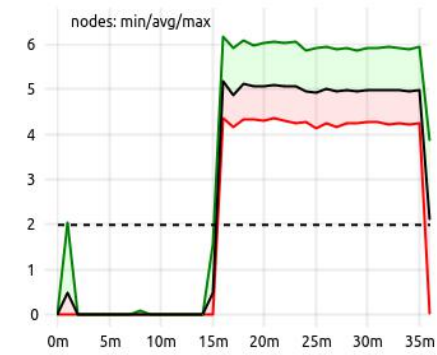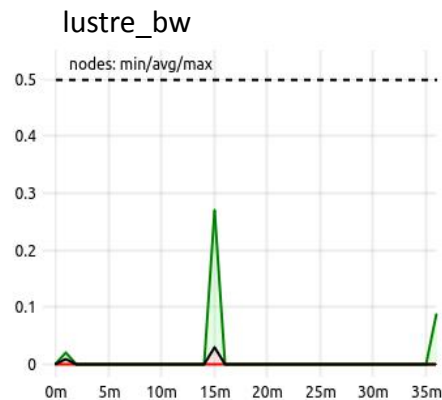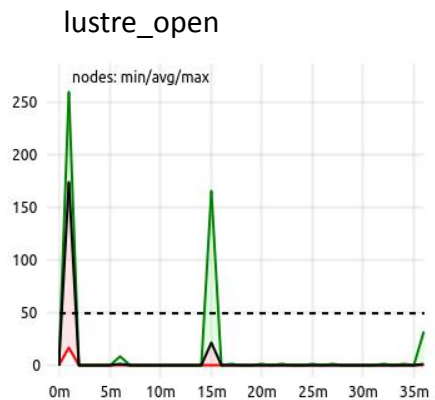
Infiniband:
100 GBit/s Full Duplex
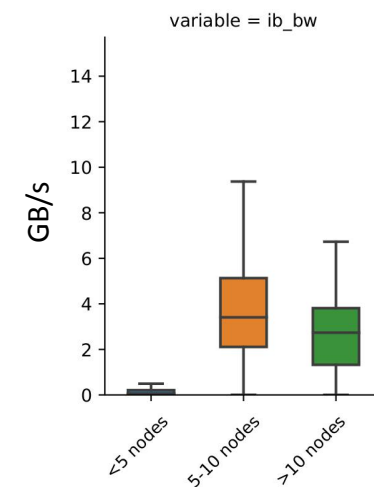
Limitations

- Node level metrics on shared/interactive

- Jobs > 2min imported every 5min

- 6 Months data retention

- No GPU FLOPS & memory bandwidth

- GPU Power disabled (aliasing)

- Statistics filter only on some metrics

Summary

- ClusterCockpit job monitoring platform

- Quick overview of running and completed jobs

- Reading job specific performance metrics

- Performance goals and limits

Federal Ministry
of Education
and Research

EE-HPC Partners:
FAU | HLRS | RWTH | HPE