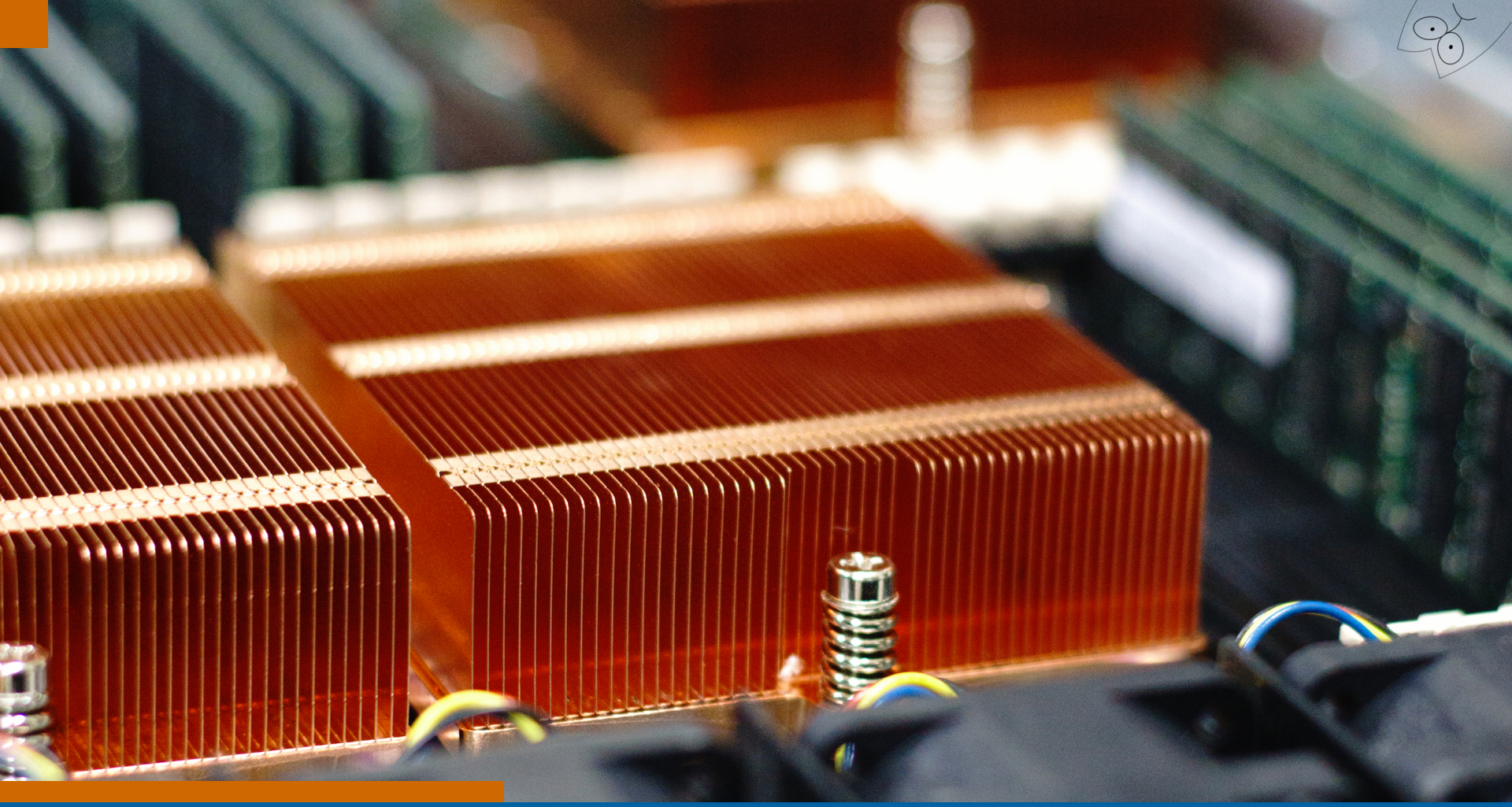


Seminar Series ‚**Hardware & Numerics**‘

4th Round – Autumn 2022

Dr. Henrik Schulz & Dr. Nina Elkina
Information Services and Computing · IT Infrastructure

Partner von DRESDEN concept exzellenz-dresden.de

HZDR

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

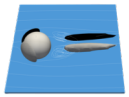
Motivation for the Seminar Series

- ✓ **Bringing together HZDR researchers**
- ✓ **Discussion about numerical methods, algorithms and computational challenges**
- ? **Improve performance of HPC calculations**
- ? **Optimal (or at least better) usage of limited resources**
- ? **Energy efficiency considerations**
- 👉 **4th round!**

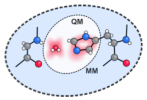
Agenda of the Seminar Series – Autumn 2022

08.11.2022 Dr. Henrik Schulz & Dr. Nina Elkina (FWCI) – lecture hall
Opening Seminar and News about HEMERA

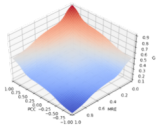
15.11.2022 Dr. Erich Focht (NEC Deutschland GmbH) – **Zoom**
The SX-Aurora Vector Engine Programming Environment



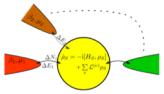
22.11.2022 Dr. Pengyu Shi (FWDC) & Dr. Nina Elkina (FWCI) – lecture hall
Fully Resolved Simulation of Wall-Bounded Bubbly Flows: Can We Do It Faster?



29.11.2022 Dr. Corey Taylor (AQEMIA) – **Zoom**
QM/MM simulation on HPCs - Introduction, Practicalities and Case Studies



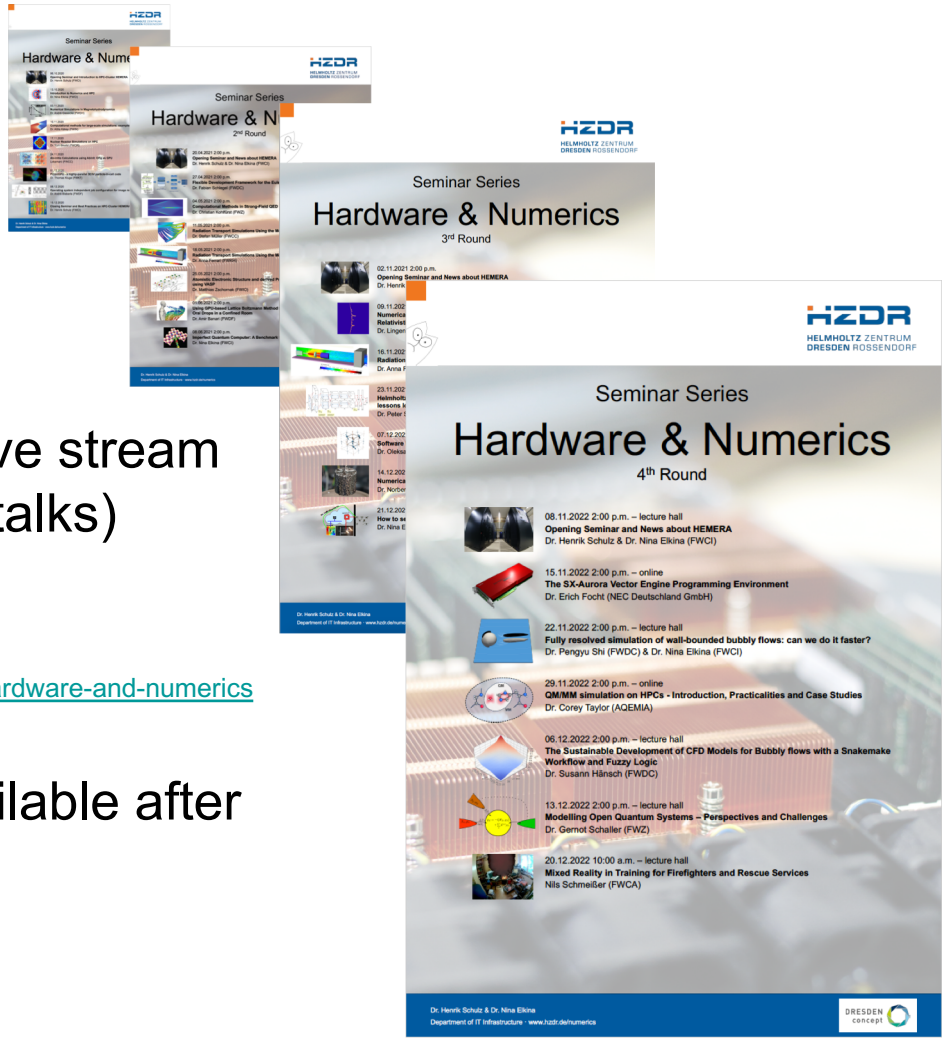
06.12.2022 Dr. Susann Hänsch (FWDC) – lecture hall
The Sustainable Development of CFD Models for Bubbly Flows with a Snakemake Workflow and Fuzzy Logic



13.12.2022 Dr. Gernot Schaller (FWZ) – lecture hall
Modelling Open Quantum Systems - Perspectives and Challenges

20.12.2022 **10:00 a.m.** Nils Schmeißer (FWCA) – lecture hall
Mixed Reality in Training for Firefighters and Rescue Services

General Information about the Seminar Series



- Seminar website:
  https://www.hzdr.de/numerics
- Online access to seminars via live stream (link will be provided before the talks)
- Chat channel during seminar:
  https://mattermost.hzdr.de/hardnum/channels/seminar-series-hardware-and-numerics
- Video recordings and slides available after talks

Seminar Series ‚**Hardware & Numerics**‘

News about **Hemera**

Dr. Henrik Schulz & Dr. Nina Elkina
Information Services and Computing · IT Infrastructure

Partner von DRESDEN concept exzellenz-dresden.de

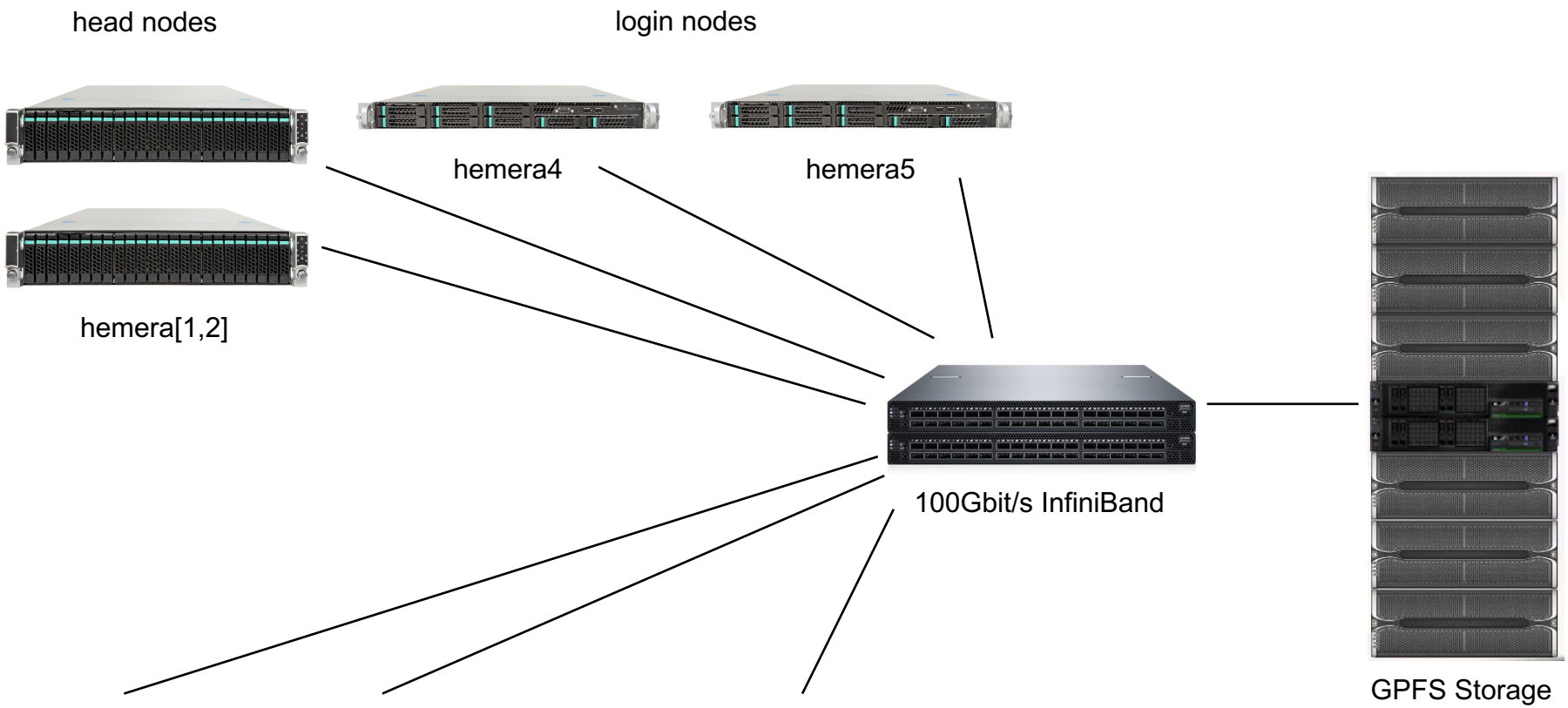
HZDR

HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

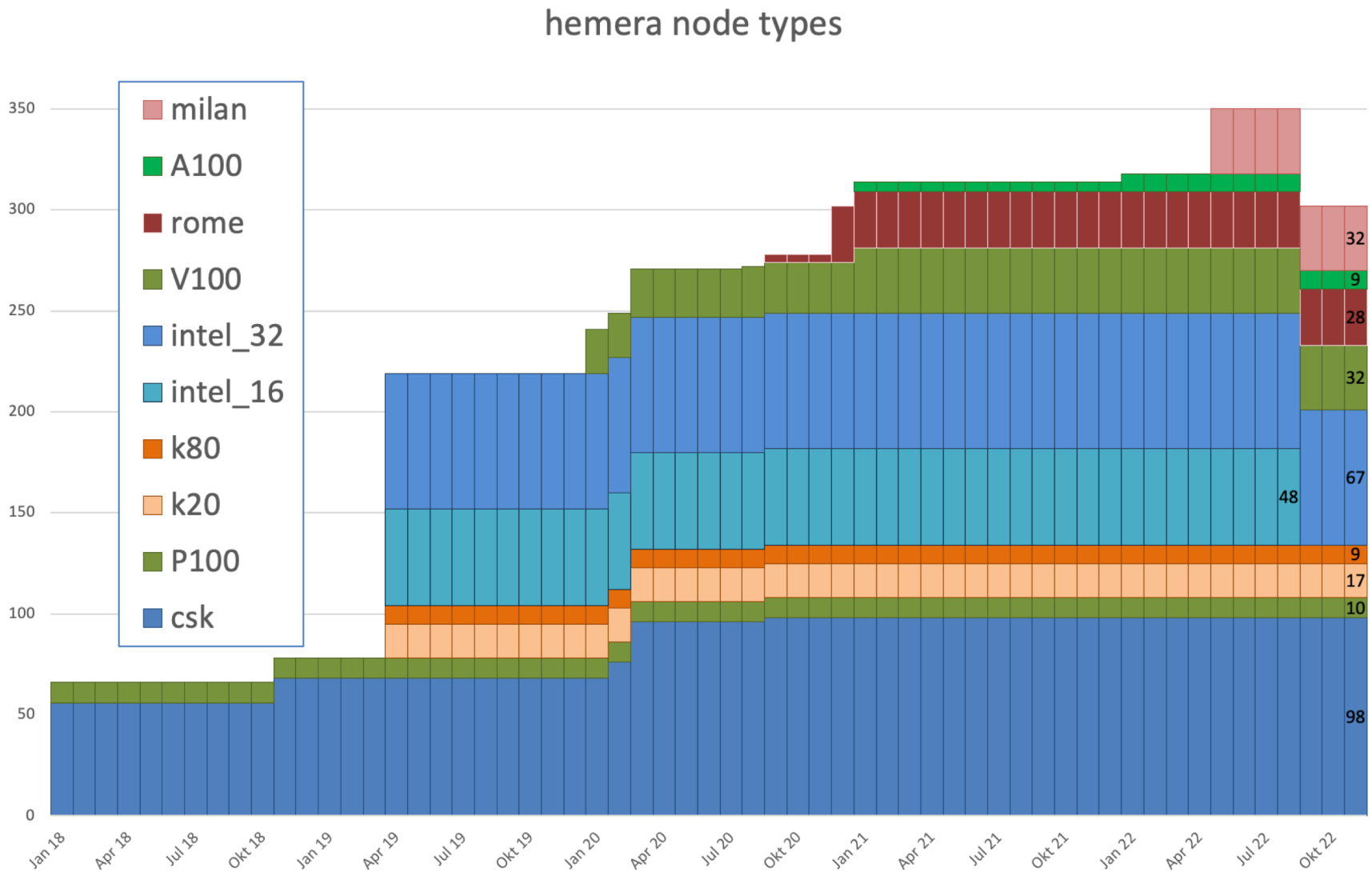
Agenda

- HPC Cluster Hemera – current status
- AMD Rome – Milan comparison
- NEC Aurora vector engine
- Information about the GSS storage system
- Usage data analysis (presented by Dr. Elkina)

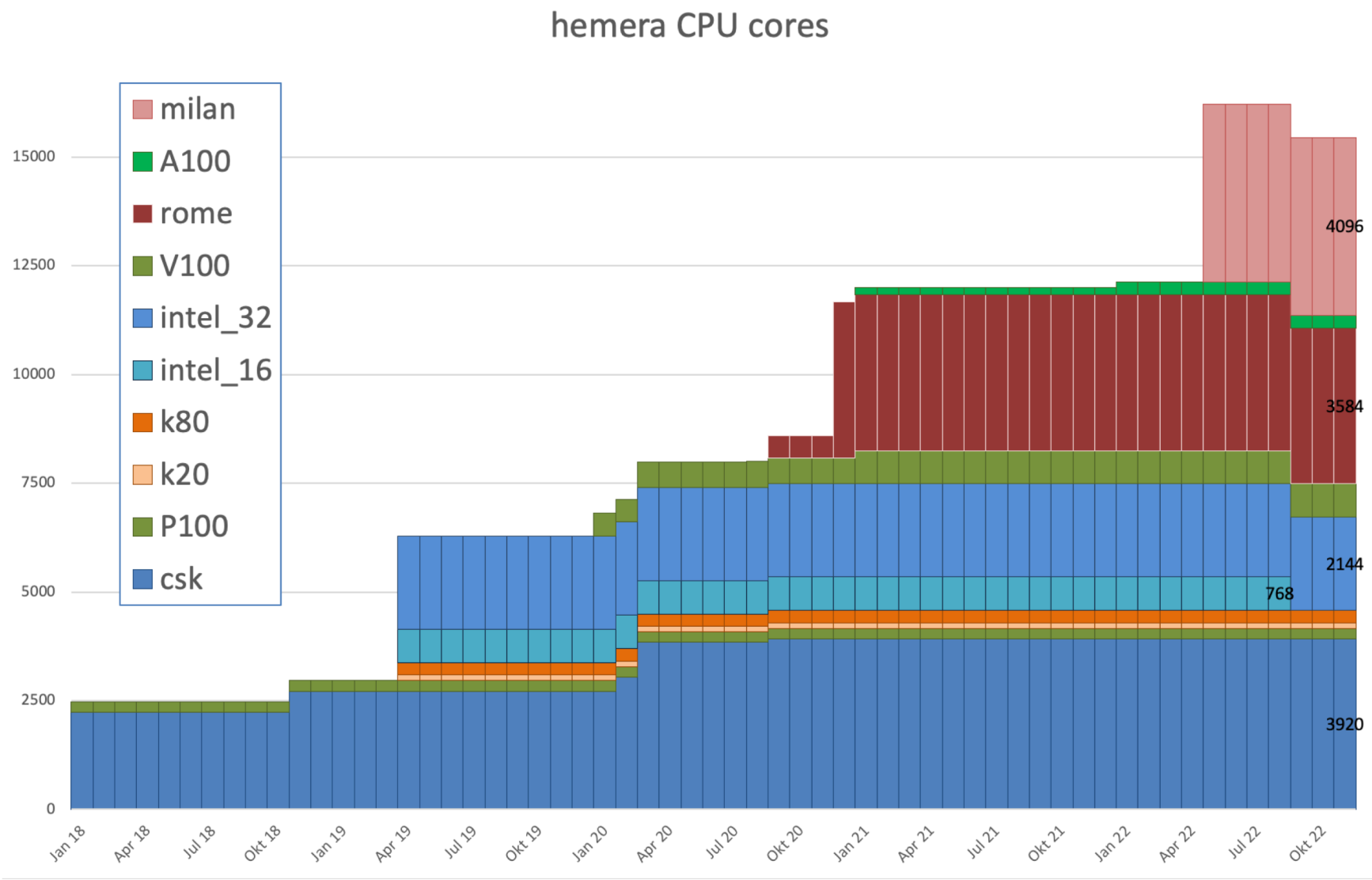
The HPC Cluster Hemera – Basic Structure



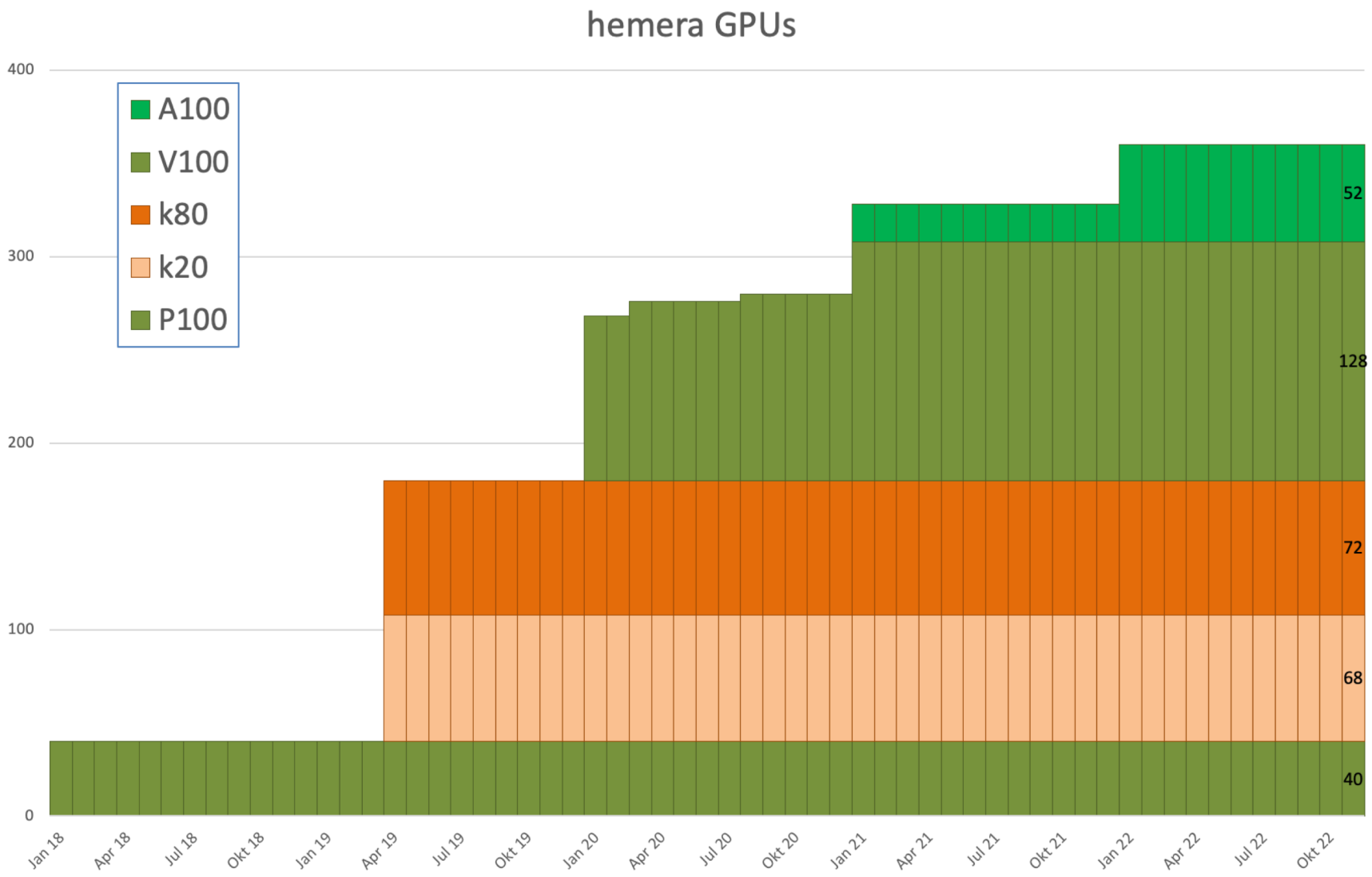
The HPC Cluster Hemera – Nodes and Partitions



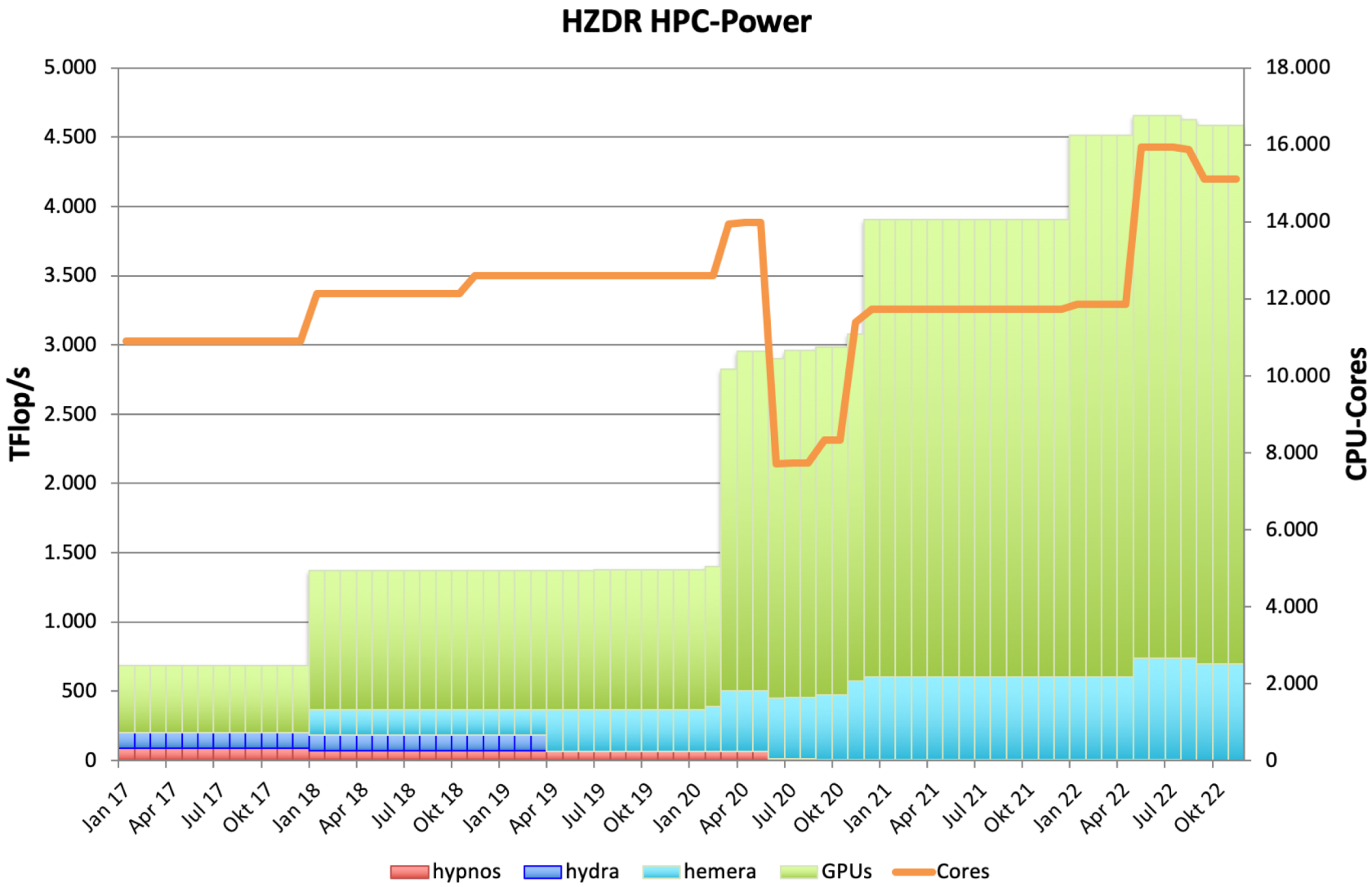
The HPC Cluster Hemera – Nodes and Partitions



The HPC Cluster Hemera – Nodes and Partitions



HPC Clusters @ HZDR



The HPC Cluster Hemera – Nodes and Partitions

Public CPU partitions:

1. defq
   - csk001-csk098 (dual-socket Intel Xeon 20-core CPU, 384 GB RAM)
2. rome
   - cro001-cro028 (dual-socket AMD Epyc2 64-core CPU, 512 GB RAM)
3. milan
   - cmi013-cro032 (dual-socket AMD Epyc3 64-core CPU, 1 TB RAM)
4. intel, intel_32
   - 67 nodes with Intel Xeon 16-core CPUs, 128 or 256 GB RAM

Limitations:

- max. walltime: 96:00:00
- max. 1024 CPUs per user
- max. 128 jobs per user

Additional partitions on these nodes: _low

- preemption partitions: low priority jobs – cancelled when nodes are needed by high priority jobs
- checkpoint/restart functionality needed
- grace time of 180s
- no walltime limitation

The HPC Cluster Hemera – Nodes and Partitions

Public GPU partitions:

5. gpu, gpu_p100, gpu_v100
   - gp001-gp010 (dual-socket Intel Xeon 12-core CPU, 384 GB RAM, 4 Nvidia P100)
   - gv022 (dual-socket Intel Xeon 12-core CPU, 384 GB RAM, 4 Nvidia V100)
   - max. walltime: 48:00:00
   - max. 32 GPUs per user

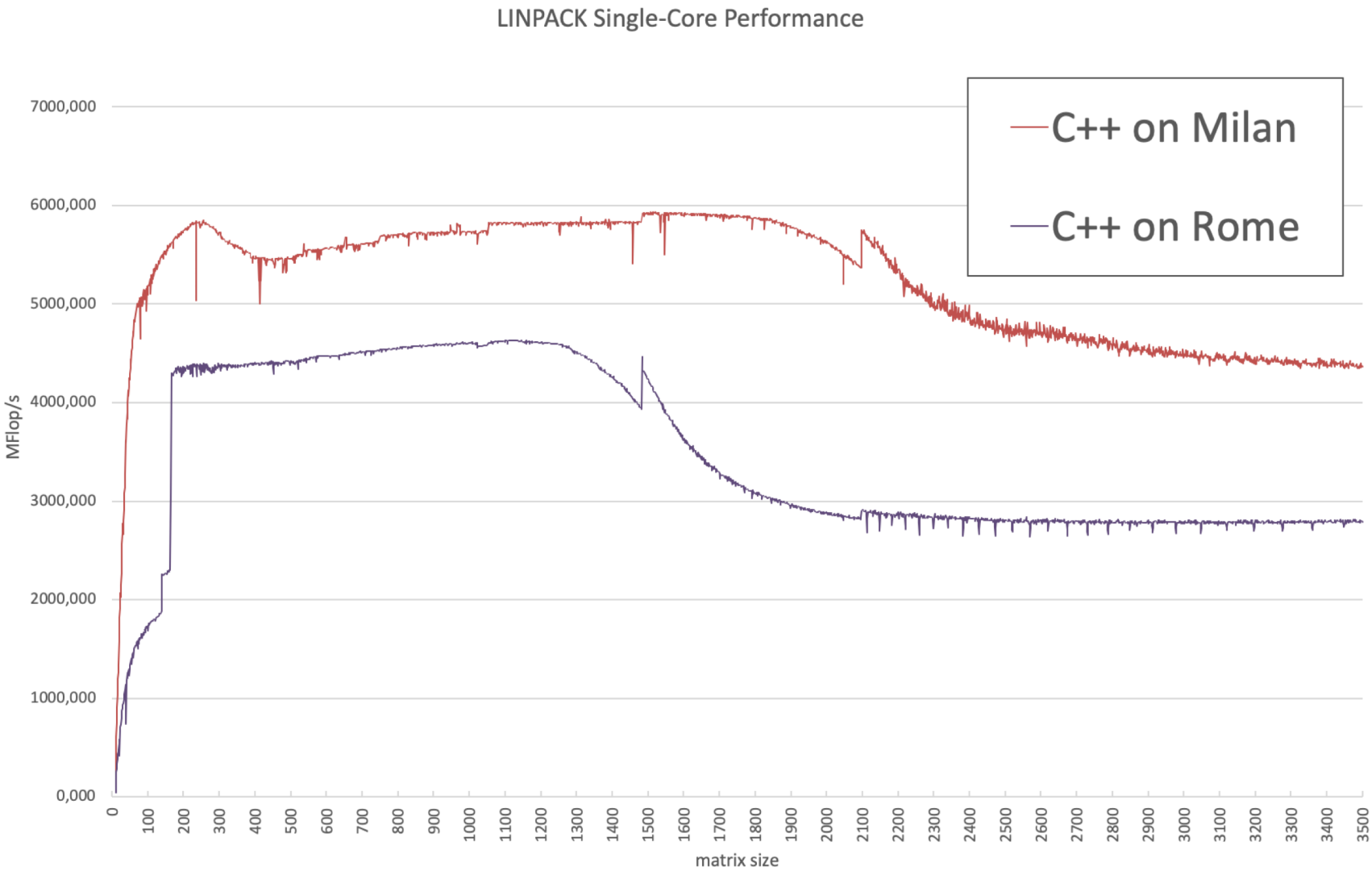
More public partitions:

6. short
   - 1 dedicated intel node
   - max. walltime: 00:15:00
   - max. 8 CPUs per user

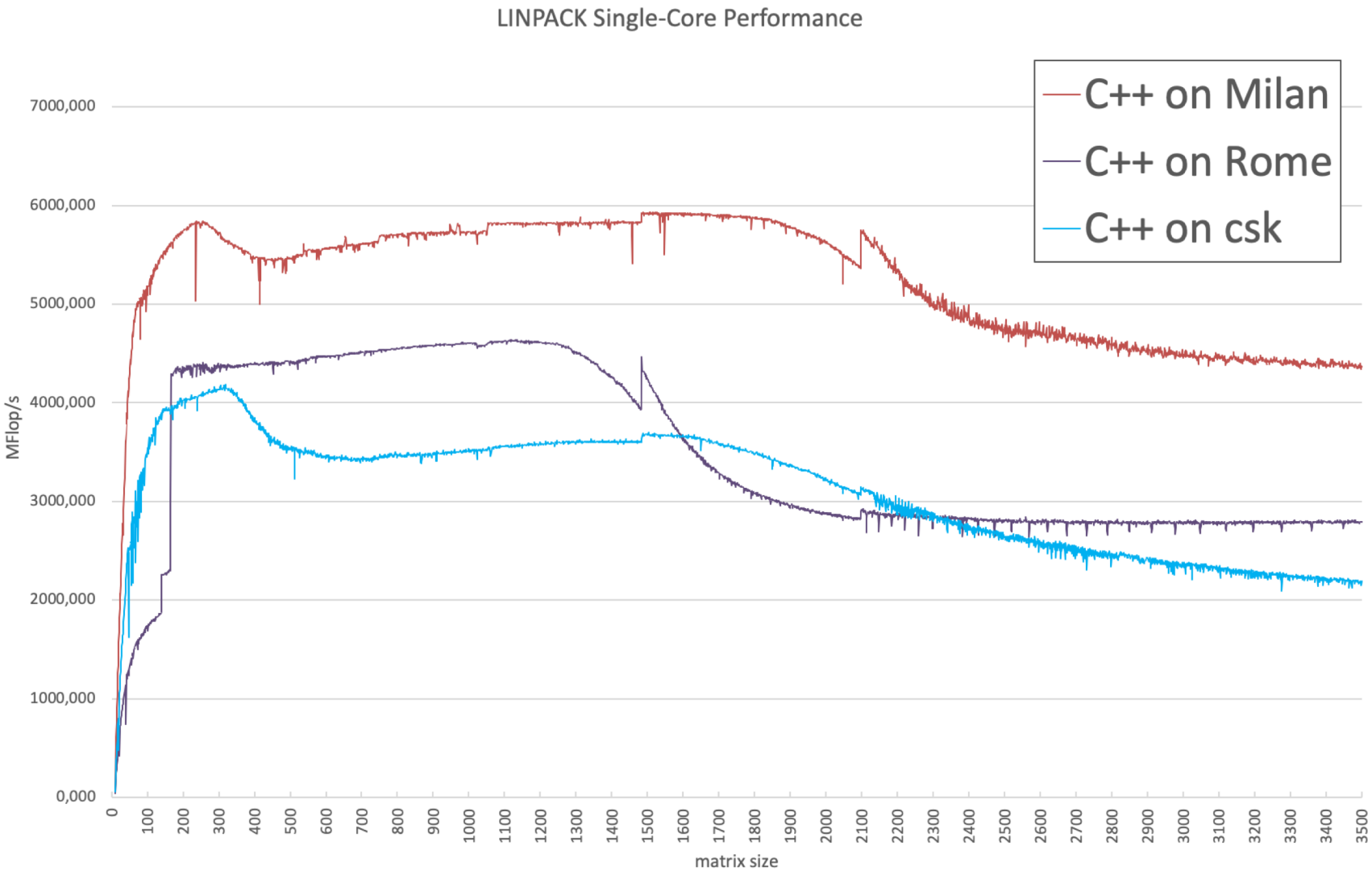
More special partitions:

- k20_low, k80_low, casus_low, a100_low
- mem768
- defq_interactive, gpu_interactive

The HPC Cluster Hemera – Nodes and Partitions



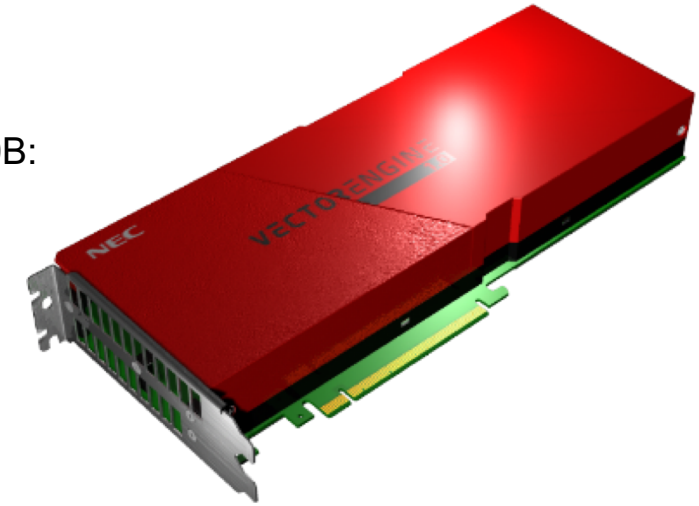
The HPC Cluster Hemera – Nodes and Partitions



The HPC Cluster Hemera – New node type

1 Node with 4 NEC Vector Engines

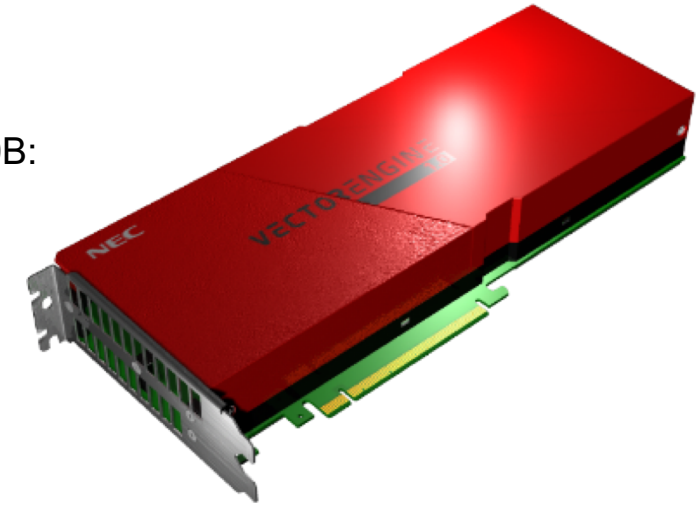
- Host contains 2 AMD Epyc 7452 CPUs and 256 GB RAM
- Each vector engine is a NEC SX-Aurora Tsubasa Type 20B:
  - 8-core processor @ 1,6 GHz
  - 16 MB cache
  - 48 GB main memory
  - 2,45 TFlop/s peak (double precision)
  - 4,9 TFlop/s peak (single precision)
  - 1,53 TB/s memory bandwidth
- NEC programming environment
  - Compiler, MPI, libraries



The HPC Cluster Hemera – New node type

1 Node with 4 NEC Vector Engines

- Host contains 2 AMD Epyc 7452 CPUs and 256 GB RAM
- Each vector engine is a NEC SX-Aurora Tsubasa Type 20B:
  - 8-core processor @ 1,6 GHz
  - 16 MB cache
  - 48 GB main memory
  - 2,45 TFlop/s peak (double precision)
  - 4,9 TFlop/s peak (single precision)
  - 1,53 TB/s memory bandwidth
- NEC programming environment
  - Compiler, MPI, libraries



Introduction to the vector architecture and programming environment will be given by Dr. Erich Focht (NEC) next week

Using Hemera

- Example job script

```
#!/bin/bash
#SBATCH --job-name=MPI_job    # name of the job
#SBATCH --partition=defq      # partition to be used
#SBATCH --time=1:00:00        # walltime (up to 96 hours)
#SBATCH --nodes=2             # number of nodes
#SBATCH --ntasks=80           # number of tasks (i.e. parallel processes) to be started
module load gcc/7.3.0
module load openmpi/2.1.2
cd $HOME/your_workdir         # path where executable and data is located
mpiexec ./simulation
```

- Submit job

```
sbatch job.sh
```

- Important information:
  - Use the minimum number of nodes possible
  - Example for partition defq:

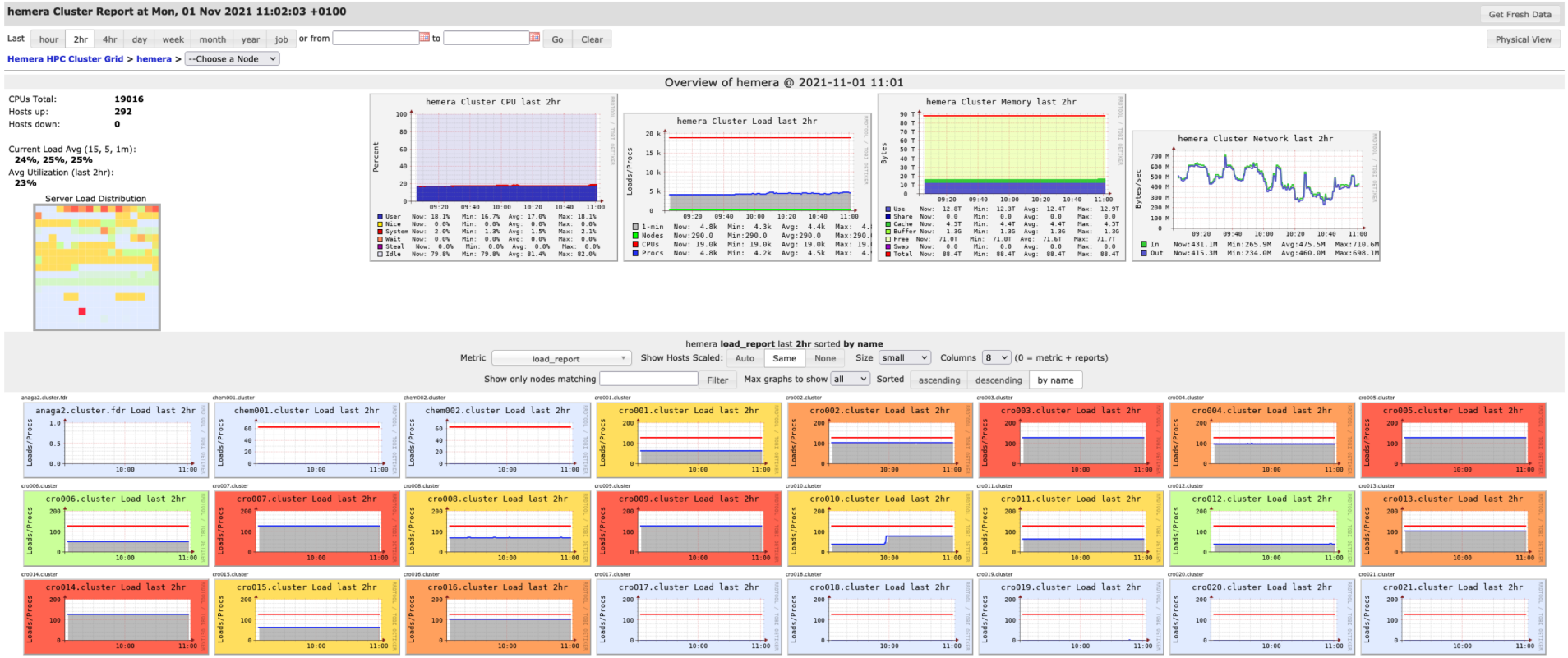
```
#SBATCH --nodes=2
#SBATCH --ntasks=80

#SBATCH --nodes=8
#SBATCH --ntasks=80
```

https://www.hzdr.de/db/Cms?pOid=59633

The HPC Cluster Hemera – Ganglia

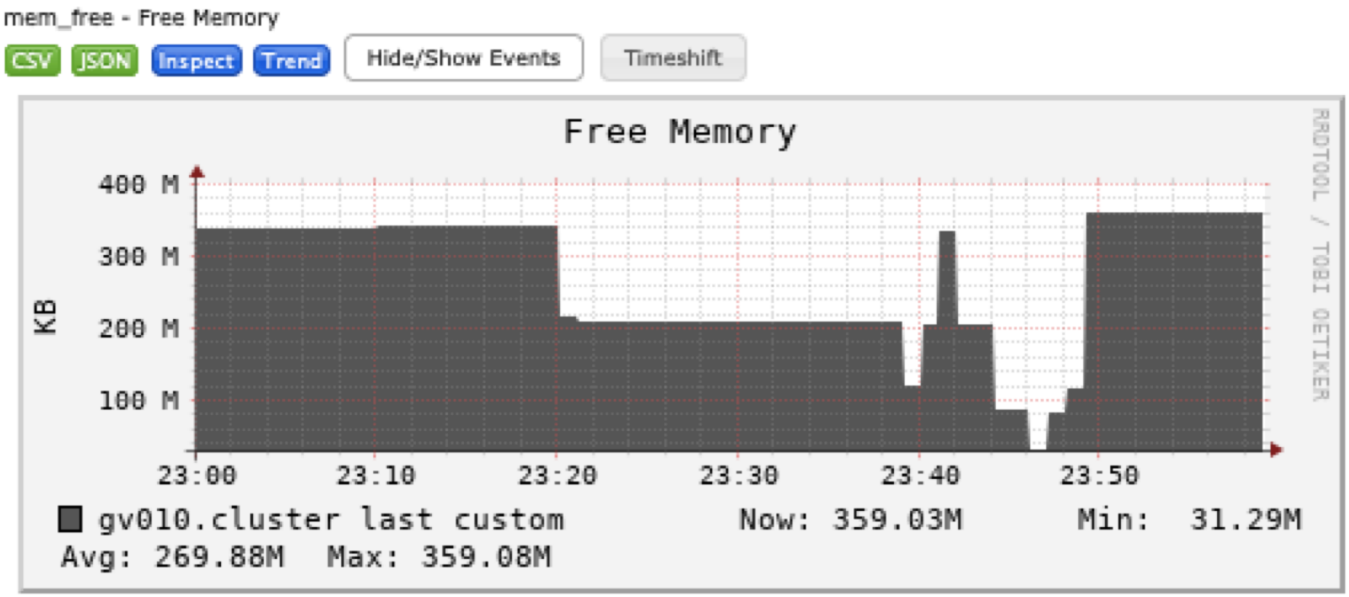
Ganglia can be used to get an overview of nodes and usage

http://hemera4/ganglia/



The HPC Cluster Hemera – Ganglia

Ganglia can be useful sometimes to find out why jobs crash…



General information about GSS

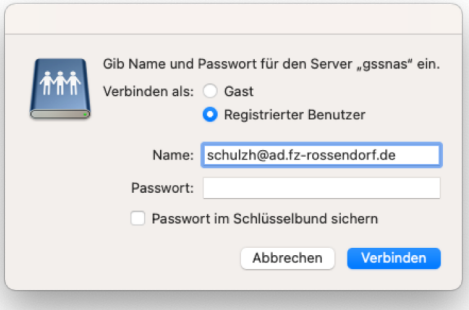
GPFS Storage System (GSS)

- 5 GxFS building blocks (by NEC):
  - dual controller NetApp storage systems
  - 480x 8TB
  - 480x 16TB
- 3 Filesystems:
  - /data - home directories (300 GB per user) – 225 TB
  - /bigdata - project directories – 5,5 PB
  - /archiv - buffer storage for tape archives – 450 TB
- Data transfer rates:
  - up to 5,8 GB/s per stream
  - up to 14 GB/s per node
  - up to 25 GB/s with many nodes
- Native directories on hemera
- Additional export via SMB and NFS:
  - gssnas
  - gssnfs
- Extension (~3 PB) already partially delivered
  - 240x 18TB

General information about GSS

Export via SMB - gssnas

- \\gssnas\bigdata
- has been updated -> workaround not needed anymore

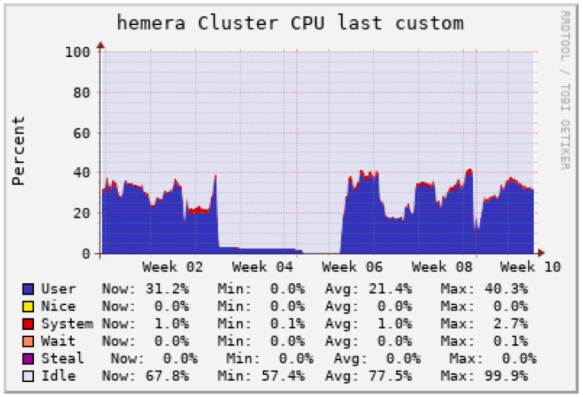


Problems with GSS in 2022:

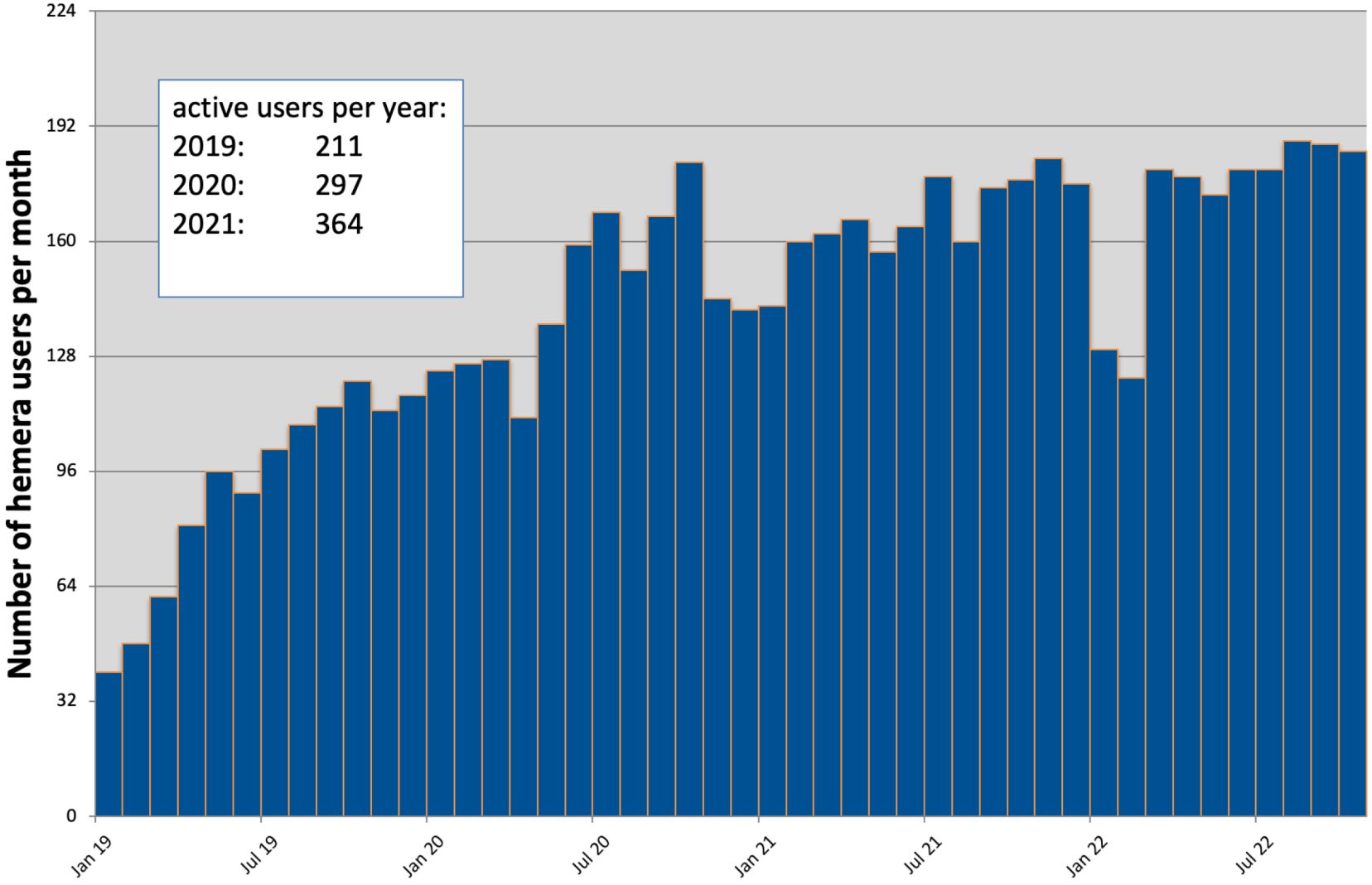
(Building blocks with 16 TB Seagate disks)

- Main problem was incompatibility between NetApp storage controllers and Seagate disks
- Media errors on disks during high workload

→ Firmware update was developed and installed
→ System is still permanently monitored by NetApp
→ Another downtime will be needed in 2022 to install extensions and rebalance hardware and filesystems

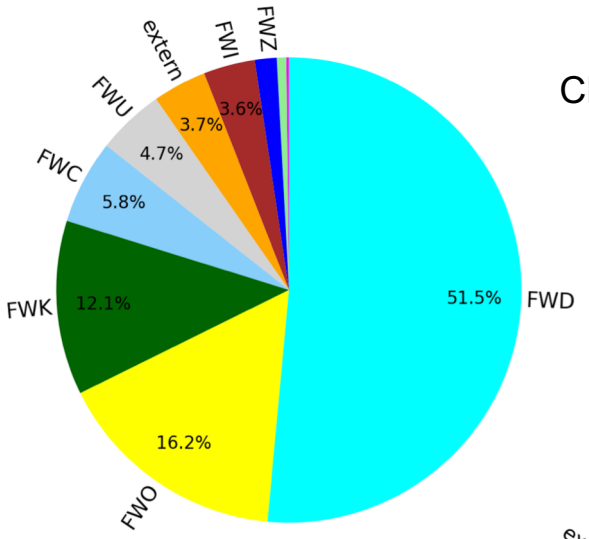


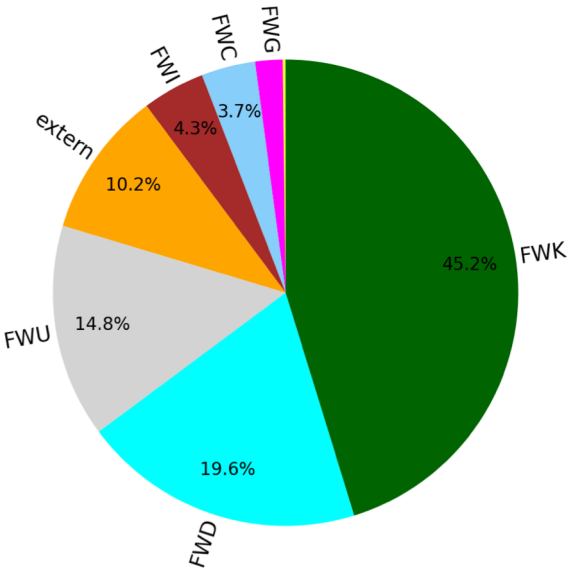
Usage data analysis



Usage data analysis

Jobs on hemera in 2022 (until Oct.): ~1.3 mio. (3% GPU)





Institute	No. of users
FWD	77
FWK	83
FWO	18
FWC	29
FWI	38
FWU	41
extern	30
FWZ	4
FWM	9
rest	2
total	**331**

Summary

- Resources in the hemera cluster are getting more heterogeneous
- Knowledge of different groups of nodes (partitions) and node architectures is needed
- Partitions have to be used according to the needs of the jobs
- Problems with GSS are solved

Thank You for Your attention!

Questions?