Metrics for performance and procurement

Monthly Storage Talks

Christian Köhler
christian.koehler@gwdg.de

April 3, 2024

hpc@gwdg.de
GWDG – Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Systems/Procurements

- NHR
- AI service center

RfPs

Evaluation criteria

- Procurement workflow

Systems/Procurements

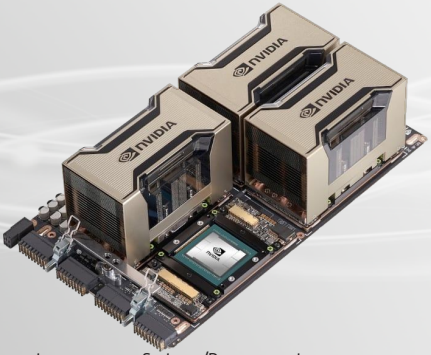
Systems/Procurements

NHR

NHR GPU system “Grete”

Overview

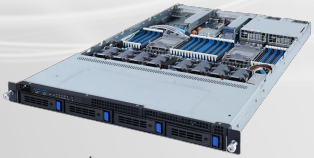
- 35 Nodes
  - 2x AMD Epyc 7513 32c, 512 GB DDR4 RAM, 2x 1 TB NVMe SSD
  - 4x NVIDIA A100 (40 GB, SXM4, 6.912/432 CUDA/Tensor cores)
  - NVLink (HGX “Redstone”), 2x 200 GBit/s InfiniBand HDR
- 2x DDN ES400NVX Storage Appliance, approx. 130 TiB flash storage



Most recent installations

NHR

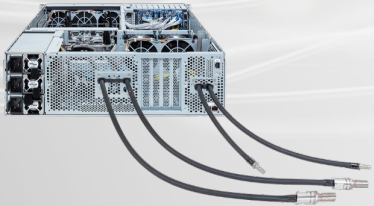
- NEC CPU cluster (replacing Emmy P1)
  - each node: 2x Intel “Sapphire Rapids” 8468 (48c)
  - 164x 256 GB, 164x 512 GB, 12x 1 TB, 2x 2 TB
  - Poweruser phase concluded
- NEC CPU Add-on 2023
  - 20x 512 GB, 16x 1 TB
  - 4 nodes per 2U chassis, each with
    - 2x Intel Sapphire Rapids 8468 (48c) CPU
    - 1x Cornelis Omni-path (100 Gbit/s) HCA



Most recent installations

NHR

- MEGWARE GPU cluster “Grete”
  - Nodes from various projects:
    34x NHR, 22x REACT, 35x AI service center, 9x hosting, 3x SCC
- MEGWARE GPU Add-on
  - 5 nodes, each with
    - 2x Intel Sapphire Rapids 8468 (48 Kerne) CPU
    - 4x NVIDIA H100 SXM5 (94 GB) GPU
    - 2x InfiniBand HDR (200 Gbit/s) HCA
  - 3U per node, DLC
  - via framework contract
  - +11 nodes for AI service center



Systems/Procurements

AI service center

Most recent installations

AI service center (KISSKI)

- Dell Cold Storage
  - Joint procurement NHR (→WORK), KISSKI, GWDG
  - Separate procurement of installation support
- VAST NVMe all-flash system (IB/OPA, 500 TB net capacity)
- Training cluster → integrated w/ Grete
- Inference cluster → Slurm+Kubernetes Setup (also for infrastructure)
- Development platform
  - Delta: NVIDIA Grace Hopper DevKit (ARM CPUs)
  - MEGWARE: Intel Habana Gaudi2 (GPU), GraphCore (IPU), Esperanto.ai (RISC-V accelerators)
- SpiNNaker
  - 4 boards*48 chips*152 ARM-cores (Neuromorphic Computing)

RfPs

Evaluation criteria

Quantitative bid evaluation

Overview

- Energy cost E

$$E = P(W \cdot 1.06 + A \cdot 1.3) \cdot T \cdot R$$

P: max. power demand; W/A water-/air-cooled part; T projected time; R electricity rate

- Performance criteria L

$$L = \sum_i \alpha_i \cdot L_i$$

α_i: individual weight; L_i: individual criteria

- Combined ranking C

$$C = L \cdot \frac{I}{B + M + E}$$

I: investment volume; B: price of the system; M: total maintenance cost

Quantitative bid evaluation

Performance criteria L_i

- Number of nodes/GPUs/net storage capacity (if applicable)
- Benchmark results
  - GROMACS, OpenFOAM, HPL, HPL-AI mixed precision
  - IO500, S3 IOPS/Bandwidth, compression effectiveness
- Infrastructure
  - Compatibility with cooling
  - No data center construction works required
- Miscellaneous criteria
  - Delivery and installation time
  - Support concept

Qualitative data are scored (according to specification), BMs usually normalized to best offer.

Evaluation criteria

Procurement workflow

Procurement workflow

- Tender process
  - Modular tender document (WIP)
  - Spreadsheet for quantitative evaluation
  - Section with data center/cooling/etc specs
- Planning
  - Regular roadmap meetings with vendors
  - User survey in advance of funding proposals
  - Aim for framework contracts (time span vs pricing)
- Organization
  - Planning with infrastructure group
  - Central storage of specs/quotes/orders/contracts
  - Clear steps in the workflow $\rightarrow$

Procurement workflow
Status tracking

- `NEW` - team notified of planned procurement
  - → create global shorthand for the procurement (metadata spreadsheed, file structure, minutes)
- `DRAFT` - tender document or mail for requesting informational quotes created
  - → present and discuss in Procurements meeting
- `INFO` - informational quotes received
  - → discuss adjustments to/feasability of evaluation criteria, iterate tender document

Procurement workflow
Status tracking

- `NOTIFY1` - Infrastructure group informed about planned system
  - include projected timeline for tender publication, evaluation week, initial order, (contractual) delivery dates
  - capacity planning (including extension options)
- `TENDER` (optional) - tender process started
  - → verify online publication of tender announcement and documents
  - notify known HPC vendors
- `QUOTES` (or bids) received
  - → evaluation according to tender criteria
- `NOTIFY2` - Infrastructure group and vendors (“Vorabinfo” §134 Abs. 1) informed about chosen system

Procurement workflow
Status tracking

- `ORDER`
  - Depending on project funds via Uni Göttingen or as GWDG
- `DELIVERY` - systems arrived on site
- `INSTALLATION` in data center completed
- `OPERATION` - customer(s) can use the systems
  → announcement to userbase (generally accessible systems) or PI (for individual procurements) with instructions on how to give their team access
- `EOL` - system is out of operation
  → decommissioning as agreed with vendor (cf. EVB-IT contract) or selling of scrap hardware
  - internal storage media have to be wiped

Monthly Storage Talks

https://hps.vi4io.org/events/2024/mst