GöHPCoffee

HPC/GPU resources and hosting/housing options

Christian Köhler
christian.koehler@gwdg.de

2. Oktober 2024

hpc@gwdg.de

HPC Systems hosted in Göttingen

Applications

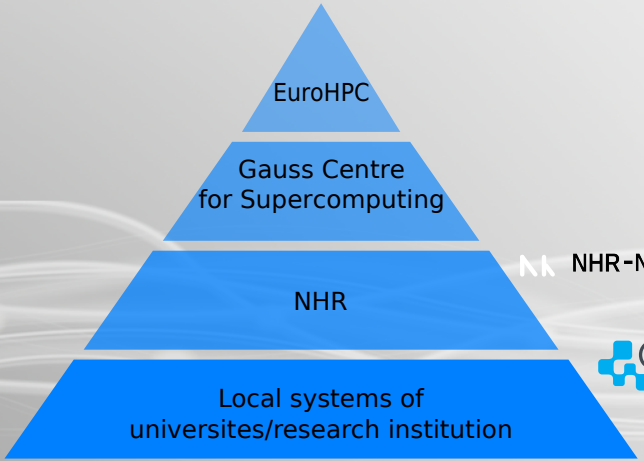
Getting Access

Sharing/Hosting/Housing

HPC Systems hosted in Göttingen

HPC landscape in Germany

T0: Europe, T1/2: Germany, T3: local



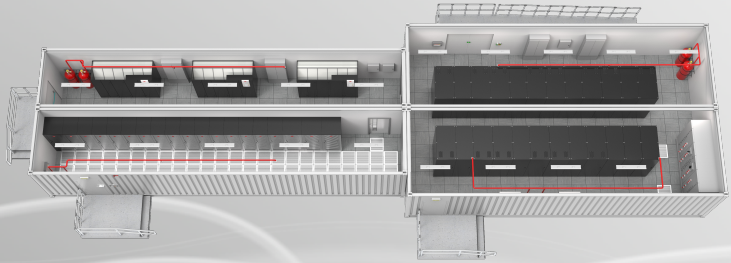
NHR-NORD@GÖTTINGEN
"Emmy", "Grete"

GWDG
Scientific
Compute
Cluster

HPC systems operated by GWDG

- Scientific Compute Cluster (SCC)
  - Replacement Procurement for the Delta cluster (d∗ nodes, Faßberg, from 2016) from NEC is part of Emmy P3 at RZGö
  - Installation in the new data center
  - Continued operation of the Atos cluster (a∗ nodes, MDC, from 2020)
- DLR system “CARO”
  - Cooperation between GWDG, University of Göttingen and German Aerospace Center (DLR)
  - System delivered and installed by NEC
  - in operation since 2022
  - GWDG: Installation, Operation and Consulting
- Göttingen Campus Institute for Dynamics of Biological Networks (CIDBN)
  - vendor SVA/DELL, technically integrated with SCC
- NHR (formerly HLRN-IV) Atos system “Emmy” + MEGWARE GPU expansion “Grete”

Modular Data Center (MDC)



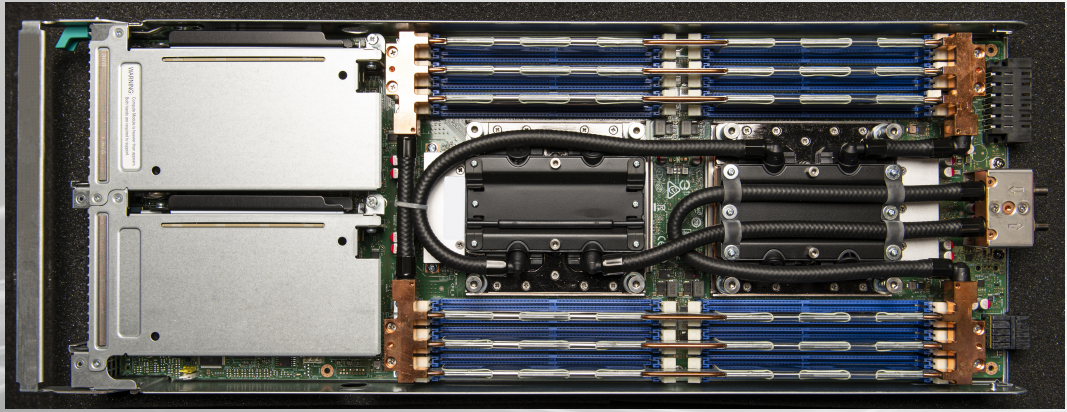
- Left section: air-cooled systems (max 19 Racks)
  "Emmy" phase 1, GWDG SCC GPU nodes, storage
- Right section: hot water-cooled systems (max 14 racks)
  "Emmy" phase 2, GWDG SCC CPU compute nodes

NHR System “Emmy”

- Phase 1 (2018) → EOL as of 2024
  - 448 standard nodes (2 Xeon Gold 6148, 2x20 CPU-Cores, 192 GB memory)
  - High Mem nodes: 16x 768 GB
- Phase 2 (2020)
  - 1004 standard nodes (2 Xeon Platinum 9242, 2x48 CPU cores, 384 GB memory)
  - High Mem nodes: 16x 768 GB, 2x 1,5 TB
  - 3 GPU nodes (2 Xeon Gold 6148, 2x20 CPU cores, 768 GB memory, 4x NVIDIA V100)
- Since 2021 operated for the NHR alliance
- Phase 3 (2023) installed in RZGö

“Emmy” P2, SCC: Intel Xeon Cascade Lake AP HPC node

Water cooling of CPUs and memory



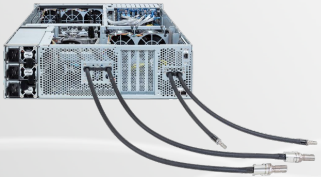
Source: GWDG

Göttinger Rechenzentrum (RZGö)

- Hosting various HPC systems
  - **“Grete”**, the GPU expansion for “Emmy”
  - “CARO” for the German Aerospace Center (DLR)
  - Max Planck Society (MPG) HPC systems (housing)
  - NHR+SCC CPU+GPU expansion 2023 (Emmy P3)
- IT staff+systems from GWDG, Uni Göttingen, UMG, MPG
- Expanded office secion, new CS lecture hall

GPU expansion „Grete“

- 34 nodes with Direct Liquid Cooling (DLC), each with
  - 2x AMD Epyc 7513 CPU (32 „Milan“ cores, Zen 3 microarch.)
  - 512 GB memory (DDR4, 3200 MHz)
  - 2x 1 TB NVMe SSD
  - 4x NVIDIA A100 GPU (SXM4, 40 GB HBM2 memory)
  - 2x Mellanox InfiniBand HCA (HDR)
- + 22x REACT, 35x AI service center, 9x hosting, 3x SCC
- GPU Add-on 2023: 5 nodes, each with
  - 2x Intel Xeon P. 8468 (48 „Sapphire Rapids “ cores)
  - 512 GB memory (DDR4, 3200 MHz)
  - 2x 1,92 TB NVMe SSD
  - 4x NVIDIA H100 GPU (SXM5, 94 GB HBM2 memory)
  - 2x Mellanox InfiniBand HCA (HDR)
- + 11x AI service center

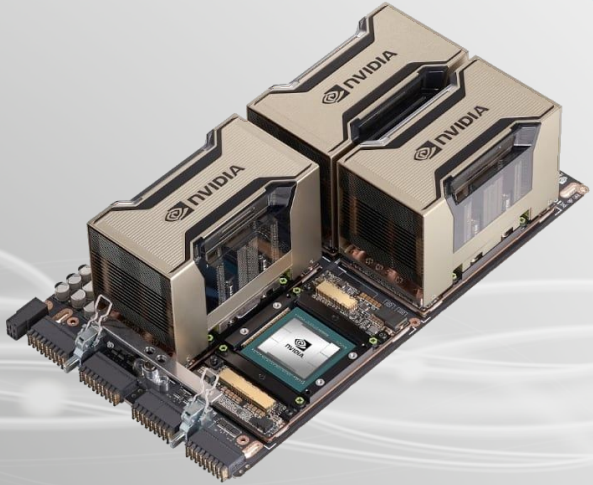


GPU System “Grete”
Specifications of the GPUs (A100 example)

- 136x A100 Tensor Core GPU (Ampere architecture)
  - GA100 GPU: 7 GPU processing clusters (GPC, share L2 cache) à $\leq$ 16 SMX
  - GPC can be split between users as Compute Instance→MIG
  - 108 SMX → 6912 CUDA cores, 432 Tensor cores
  - 19,49/9,746/155,92 TFLOPs (SP/DP/Tensor)
  - 40 GB VRAM (HBM2e)
  - SXM4 form factor, 4 GPUs on “Redstone” baseboard
  - NVLINK full mesh, 800 GB/s bidirectional bandwidth per GPU
- GPUS can be requested individually in the batch system (Slurm)

GPU example: NVIDIA “Redstone” Baseboard

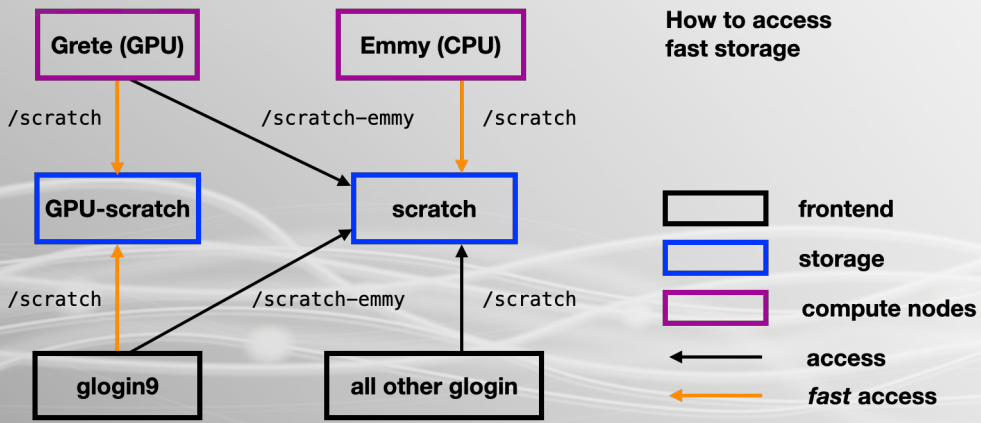
4x A100 SXM4



Source: NVIDIA

GPU System “Grete”

Frontend Nodes

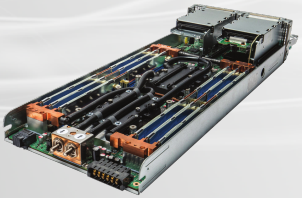


Most recent installations
NHR

- NEC CPU cluster Emmy Phase 3 (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
- SCC partition: 44x 512 GB, 4x 1 TB, 1x 2 TB

Scientific Compute Cluster
Replacement procurement 2020 - Overview

- 96 standard nodes (2 Xeon Platinum 9242, 2x48 CPU cores, 384 GB memory)
  → cf. “Emmy” Phase 2 standard nodes
- 2 GPU nodes with 8 Tesla V100/32
  →2 Xeon Gold 6252, 2x24 CPU cores, 384 GB memory
- 12 GPU nodes with 4 Quadro RTX5000
  →2 Xeon Gold 6242, 2x16 CPU cores, 192 GB memory
- 100 GBit/s Intel Omni-Path interconnect
- Storage: 2,15 PB /scratch
  → BeeGFS, with a 100 TB SSD part



- 48x Quadro RTX5000 (Turing architecture)
  - TU104 GPU: 48 SMX → 3072 CUDA cores, 384 Tensor cores
  - 11,2/89,2 TFLOPs (SP/Tensor)
  - 16 GB VRAM (GDDR6)
  - + 48 RT cores for raytracing acceleration=handing off lots of intersection tests from CUDA cores/shaders)
- 16x Tesla V100/32 (PCIe version) (Volta architecture)
  - GV100 GPU: 80 SMX → 5120 CUDA cores, 640 Tensor cores
  - 14/7/112 TFLOPs (SP/DP/Tensor)
  - 32 GB VRAM (HBM2)
- GPUS can be requested individually in the batch system (Slurm)

GPU example: NVIDIA Quadro RTX 5000

Source: NVIDIA

Future Technology Platform

- Experimental applications on an alternative microarchitecture (ARM) and security/storage offloading (DPUs)
- First system: NVIDIA ARM DevKits
  - 1x Ampere Altra Q80-30 (80 cores, ARM)
  - 512 GB memory (DDR4, 3200 MHz)
  - 2x NVIDIA A100 GPU (SXM4, 40 GB HBM2 memory)
  - 2x NVIDIA BlueField-2 E-Series DPU (200 GbE/HDR, 16 GB memory)
- New procurements (via AI service center)
  - delivered, now network, configuration, OS and SW setup
    - GraphCore
    - Intel Habana Gaudi 2
    - NVIDIA Grace Hopper (ARM+H100)
    - Esperanto.AI (RISC-V)
  - SpiNNcloud (ARM) → ETA December, test system (loan TUD installed)
    `https://info.gwdg.de/news/gwdg-wird-mit-neuromorpher-spinnaker-hardware-ausgestattet/`

Applications

Typical Usage Scenarios

AI Research

- Labs with many students doing small-scale experiments
- Mid-scale development/experiments by researchers or students
- Large-scale/scalable execution of developed pipelines

HPC Research

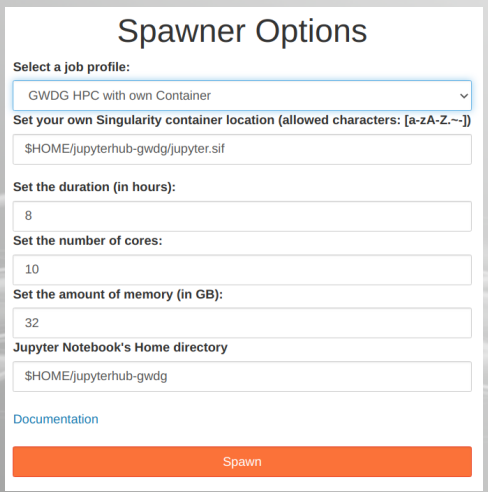
- Small-scale / single-node testing of an application
- Labs with many students learning to use HPC
- Scalable/Multi-node execution of existing/developed application

GPU Usage Modes

- **Asynchronous batch jobs**
  - submitted jobs are executed later
  - depending on load, individual priority (cf. FairShare)
  - once running resources are granted exclusively
  - most resources, main form of GPU usage
- **Interactive batch jobs**
  - resources are requested in batch mode
  - live usage once jobs starts
  - exclusive usage, potentially long waiting time

  We partition GPUs for maximum sharing (into up to 7 "devices")
- **Shared resources**
  - partition with no exlusive usage (overprovisioning)
  - no waiting time, but no guaranteed performance
  - for testing and small-scale experiments
  - usage via Jupyter HPC $\rightarrow$

Data Analytics Applications
Jupyter on SCC



- jupyter-hpc.gwdg.de spawns Jupyter on HPC
- supports IPython Parallel
- Users can choose resources, start individual Singularity Container
- Support for Jupyter on (shared) GPU nodes

Data Analytics Applications

Apache Spark / Tensorflow

- Apache Spark
  - Spark clusters can be set up automatically
  - interaktive sessions (e.g. Scala)
  - Monitor cluster status via web interface
- TensorFlow
  - Integration in users' Python environment
  - Using GPU nodes with CUDA possbile

Overview & Documentation of HPC applications: https://docs.hpc.gwdg.de

Module system

- Application domains & software
  - Quantum chemistry, Molecular dynamics
  - Bioinformatics, Genomics, Evolutionary Biology
  - Astrophysics, Cosmology, Numerical Fluid Dynamics
  - Numerical Software: MATLAB, Maple, Mathematica
  - Machine Learning
  - Medical Imaging (e.g. MRI)
  - Data Analytics: Python, R, Spark, ...
- Module system `Lmod` → software catalogue `module avail`

Getting Access

Usage Scenarios Tier 2/3

- **Scientific Compute Cluster (SCC)** (Tier 3) for UGOE/MPG
  - Covering basic usage, Throughput-Computing
  - Consulting for the transition to T2, Scaling tests
  - Requirements
    - GWDG Full Account (Nutzer UGOE/MPG)
    - Thesis projects: Apply for an account via supervisor's institute
    - Courses: temporarily usable guest accounts
- **NHR systems "Emmy"/"Grete"** (Tier 2)
  for Universities and Research Institutions in the NHR alliance
  - Proposal preparation, Software setup
  - Scientific review of compute time projects

Getting access to the HPC systems

- **Scientific Compute Cluster (SCC)**
  - Account activation: request via hpc@gwdg.de
  - Job prioritization via FairShare
- **NHR systems “Emmy”/“Grete”**
  - Apply for a “Starter” project https://jards.nhr-verein.de/
  - Starting from 300.000 Coreh per quarter
    - Project application for further compute time “Normal”/“Large”
    - Review by scientific board (+techn. review by HPC sites)
    - Scientific review can be omitted , if project has been granted by BMBF/DFG/EU/GCS/NHR/... (Whitelisting) and not “Large”
- Get login account, add project members via https://hpcproject.gwdg.de

Sharing/Hosting/Housing

Integration of own systems

- **Housing** (administration by users) includes:
  - Rack space
  - Power
  - Cooling
  - Networking connectivity
- In addition, with **Hosting** GWDG takes care of:
  - Provisioning Node images
  - optional integration with the batch system (FairShare compensation or exclusive usage)
- How to include **own procurements**
  - Stating your requirements, initial consulting
  - Deciding on housing/hosting, batch system integration
  - GWDG: Procurement, order, preparing infrastructure, installation
  - further coordination: granting user access, project directories

Further Resources

Where to get more information?

- **HPC documentatin** `https://docs.hpc.gwdg.de` shows you information about
  - Basic access
  - Hardware overview
  - How to submit jobs
  - Advice on most common applications
- **Support** mail address: `hpc@gwdg.de` / `nhr-support@gwdg.de` / `kisski-support@gwdg.de`
- General GWDG support address: `support@gwdg.de`
- Rocket.Chat channel `#hpc-users` (N.B.: Matrix)

Courses

- We offer **courses** on SCC usage, programming with MPI, OpenMP, CUDA, Python, Research Data Management and more
  → **GWDG Academy** `https://www.gwdg.de/academy`
- HPC intro course **“Using the Scientific Compute Cluster”**
  - usually online via BigBlueButton

Q&A

All presentations in the GöHPCoffee series:

https://docs.gwdg.de/doku.php?id=en:services:application_services:high_performance_computing:hpc_coffee