National High Performance Computing for Science

HASCO Summer School 2024

Christian Köhler
christian.koehler@gwdg.de

August 5, 2024

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

NHR-NORD@GÖTTINGEN

The NHR alliance

HPC at GWDG

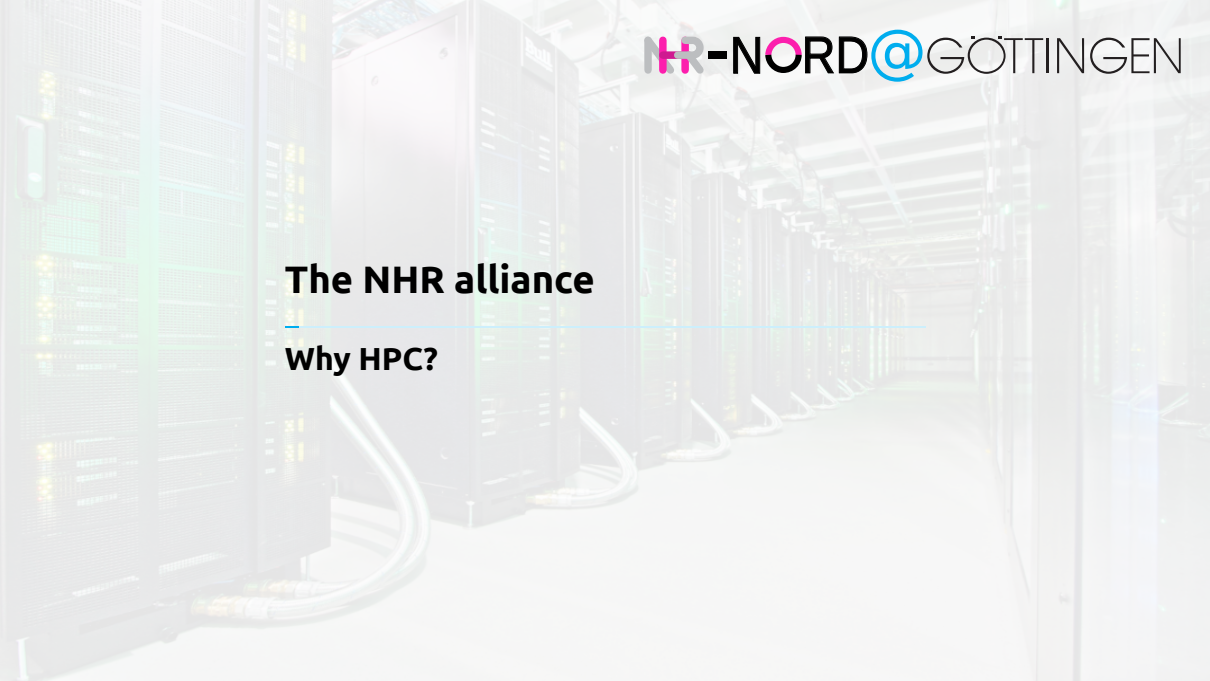
Excursion: Using Graphics Cards for Scientific Computing

Application Software

Getting access

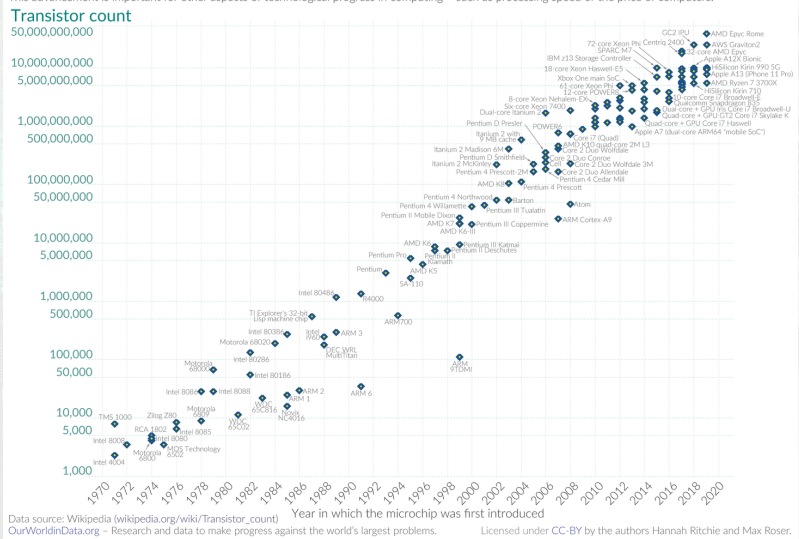
Further resources

The NHR alliance



The NHR alliance

Why HPC?



https://de.wikipedia.org/wiki/Mooresches_Gesetz#/media/Datei:Moore's_Law_Transistor_Count_1970-2020.png
Max Roser, Hannah Ritchie - https://ourworldindata.org/uploads/2020/11/Transistor-Count-over-time.png [CC BY 4.0)]

Comparison: CPU/GPU FP Operations

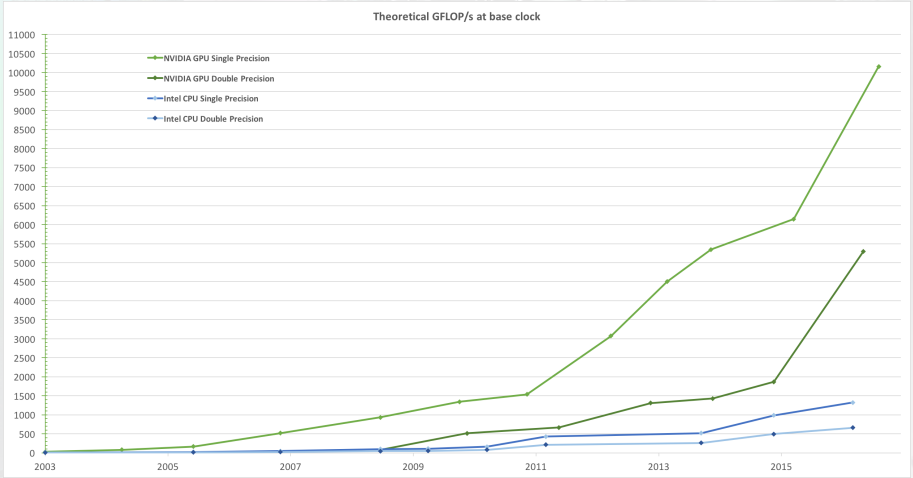
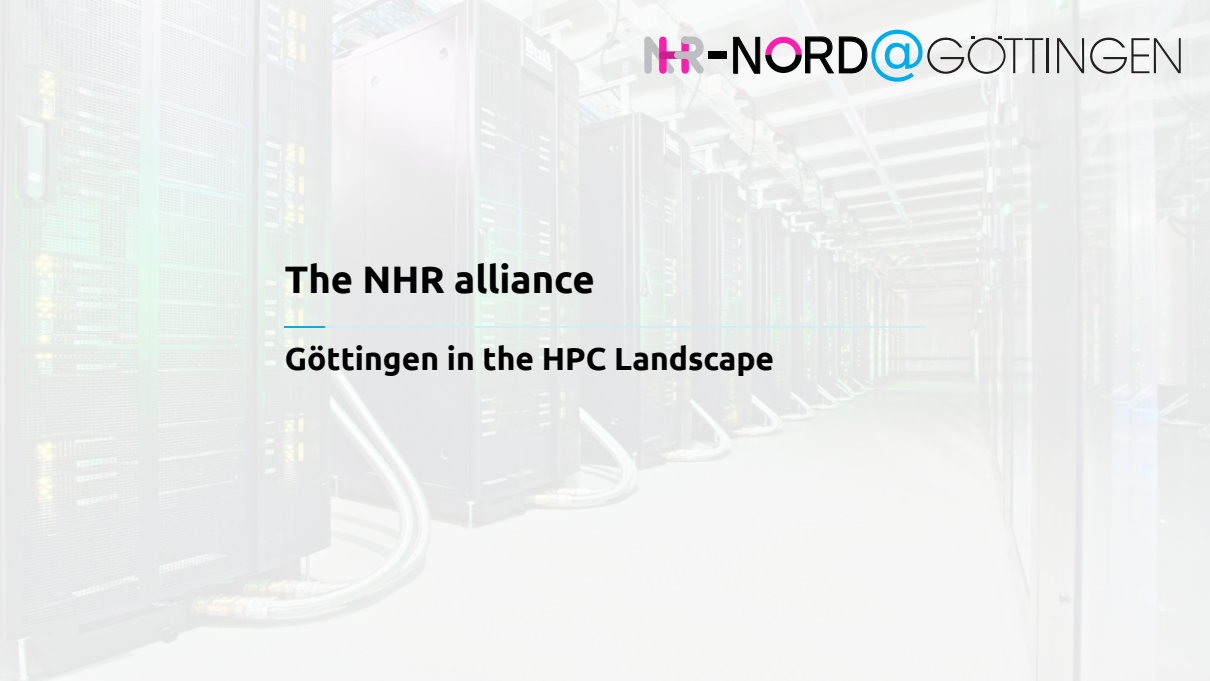


Image source: nVidia - CUDA C Programming Guide



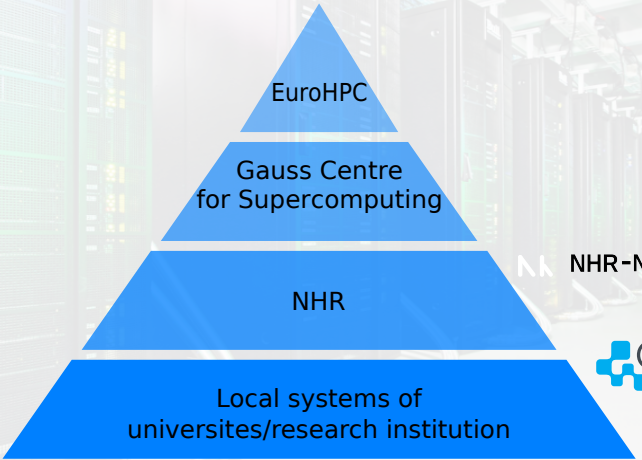
The NHR alliance

Göttingen in the HPC Landscape

HPC landscape in Germany

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

NHR-NORD@GÖTTINGEN



NHR-NORD@GÖTTINGEN
"Emmy", "Grete"

GWDG
Scientific
Compute
Cluster

HLRN background

NHR-NORD@GÖTTINGEN

- HLRN (North German Supercomputing Alliance) exists since 2001
- originally joint effort of Berlin, Brandenburg, Hamburg, Mecklenburg-Vorpommern, Lower Saxony, Schleswig Holstein (Brandenburg joined in 2012) + federal funding
- HLRN-I to HLRN-III operated by RRZN Hannover (now LUIS) and Zuse Institute Berlin (ZIB)
- HLRN-IV started operation in 2018, operated by UGOE/GWDG and ZIB
- 2020: application for NHR funding by both sites
  - originally 8 centers funded in total
  - joined by NHR@SW

NHR centers

NHR-NORD@GÖTTINGEN

- **NHR4CES@RWTH** - IT Center - RWTH Aachen
- **NHR4CES@TUDa** - Hochschulrechenzentrum (HRZ) - Technische Universität Darmstadt
- **ZIH** - Zentrum für Informationsdienste und Hochleistungsrechnen - Technische Universität Dresden
- **NHR@FAU** - Regionales Rechenzentrum Erlangen - Universität Erlangen-Nürnberg
- **NHR@KIT** - Steinbuch Centre for Computing (SCC) - Karlsruher Institut für Technologie
- **PC2** - Paderborn Center for Parallel Computing - Universität Paderborn

NHR centers

NHR-NORD@GÖTTINGEN

- **NHR-Nord@Göttingen** - Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen - Universität Göttingen
- **NHR@ZIB** - Zuse-Institut Berlin - Berlin University Alliance
- **NHR South-West**
  - Goethe-Universität Frankfurt
  - Rheinland-Pfälzische Technische Universität Kaiserslautern-Landau (RPTU)
  - Johannes Gutenberg-Universität Mainz
  - Universität des Saarlandes

Thematic focus of NHR@ZIB, NHR@Göttingen

NORD@GÖTTINGEN

- Applications
  - Life Science (RELION, BART, AMIRA, GROMACS, NAMD, ...)
  - CFD (OpenFoam, TAU, Ansys, ...)
  - Earth System Sciences (PALM, FESOM, ...)
  - Digital Humanities (Python, R, LLMs, ...)
  - Big Data and AI (Spark, TensorFlow, PyTorch, Jupyter, ...)
- Methods, Hardware, Operations
  - Data and Workflow Management
  - Virtualization and Runtime Environments
- Tools
  - ProfiT-HPC Performance monitoring
  - HPC-API for CI/CD and Workflow applications
- Teaching & Training
  - GWDG Academy, HLRN Courses
  - HPC Introduction, MPI, CUDA, HPDA, Containers in HPC, Object Storage

Competence network

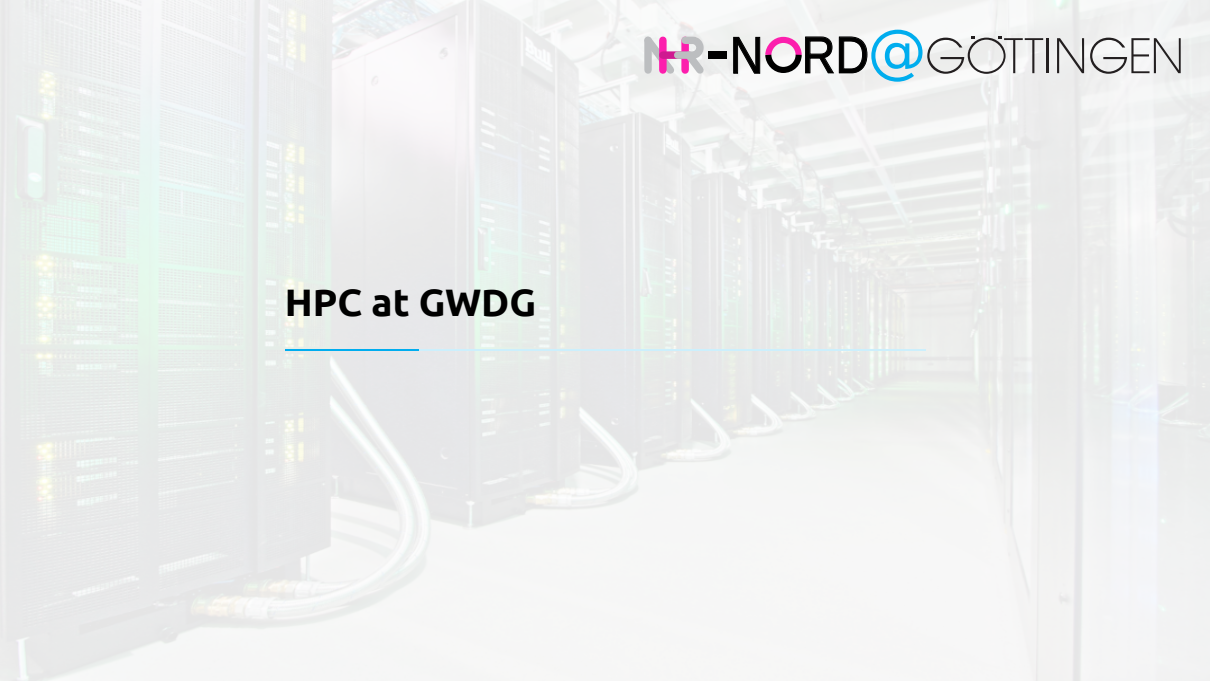
NHR-NORD@GÖTTINGEN

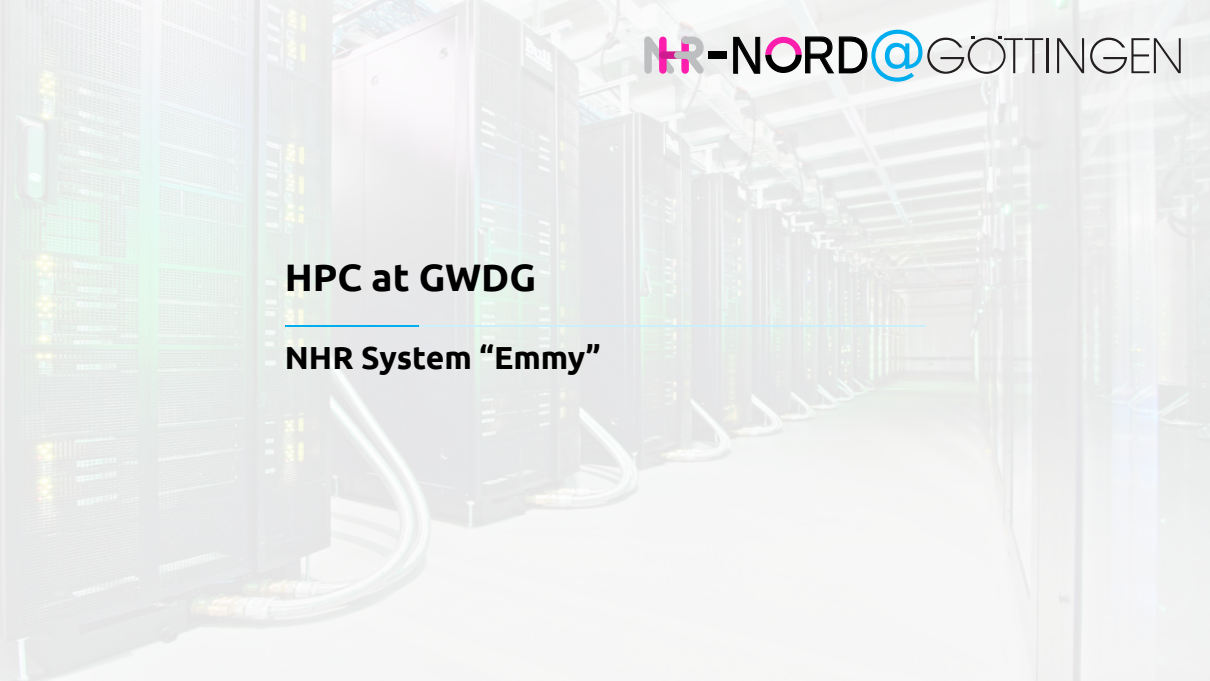
- Research Center for Many-core HPC IPCC in Berlin
- Berlin Big Data Center BBDC
- Berliner Zentrum für Maschinelles Lernen BMZL
- Göttinger HPC Verbund GöHPC
- Campus Institute Data Science (CIDAS) in Göttingen
- Kompetenzzentrum für Höchstleistungsrechnen Bremen BremHLR
- Computational Sciences Center CSC in Kiel
- Lothar Collatz Center in Hamburg
- $\rightarrow$ tier 3 sites in the 7 HLRN states

Consultants network

NHR-NORD@GÖTTINGEN

- Consultants (domain specific)
  - Support with system usage
  - Applying for compute projects (in particular estimating demand)
  - Bi-annual workshops: procurements (tier 2 und 3), system security, operations, projects, ...
- Local consultants (in part also active as domain spec. consultants)
  - Support with the transition tier 3 $\rightarrow$ tier 2
  - Initial account application


HPC at GWDG



HPC at GWDG

NHR System “Emmy”

NHR System “Emmy”

NHR-NORD@GÖTTINGEN

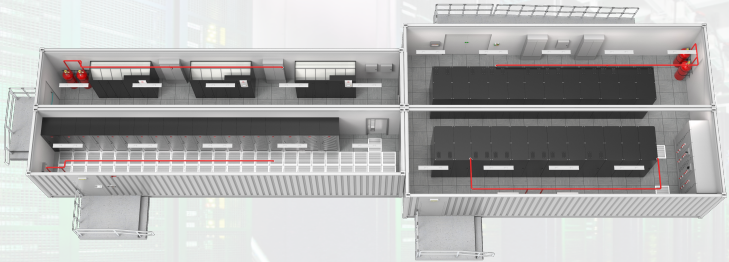
- Phase 1 (Atos, 2018) → EOL
  - 448 standard nodes (2 Xeon Gold 6148, 2x20 CPU-Cores, 192 GB memory)
  - High Mem nodes: 16x 768 GB
  - Storage: 340 TiB `HOME` (GPFS), 8.1 PiB `WORK` (Lustre), 120 TiB `PERM` (Tape)
- Phase 2 (Atos, 2020)
  - 96 standard nodes (2 Xeon Platinum 9242, 2x48 CPU cores, 384 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

NHR System “Emmy”
Phase 3

NHR-NORD@GÖTTINGEN

- Phase 3 CPU cluster (NEC, 2022, successor Emmy P1)
  - 164x 256 GB, 164x 512 GB, 12x 1 TB, 2x 2 TB
  - 4 nodes per 2U chassis, each with
    - 2x Intel “Sapphire Rapids” 8468 (48 cores) per node
    - 1x Cornelis Omni-path (100 Gbit/s) HCA
- CPU Add-on 2023 (NEC, 2023)
  - 20x 512 GB, 16x 1 TB

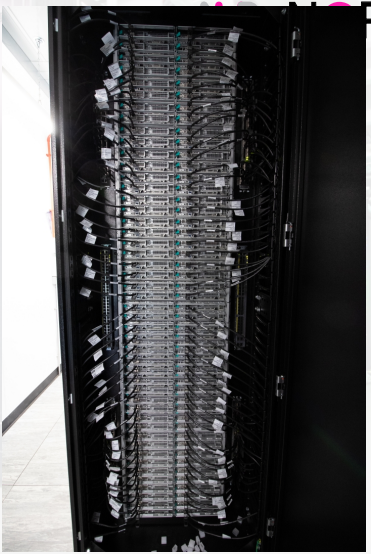
Mobile Data Center (MDC)



- 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

“Emmy” Phase 1
compute nodes

Source: GWDG

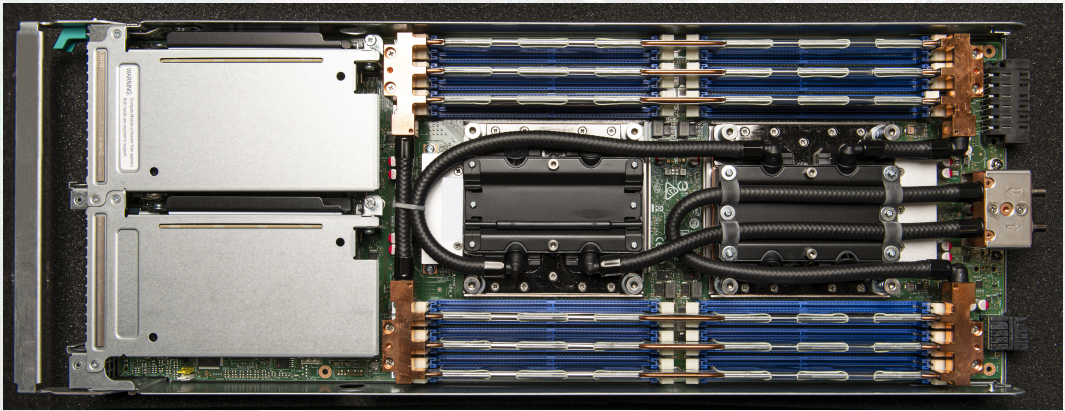


“Emmy” Phase 2
front/back

Source: GWDG

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

Water cooling of CPUs and memory



Source: GWDG

“Emmy” Storage

controllers+JBODS

Source: GWDG

MDC cooling
primary loop

Source: GWDG

MDC cooling
heat exchangers

Source: GWDG

MDC
outside view

Source: GWDG

NHR-NORD@GÖTTINGEN

HPC at GWDG

NHR GPU System “Grete”

Göttinger Rechenzentrum (RZGö)

NHR-NORD@GÖTTINGEN

- Hosting various HPC systems
  - **“Grete”**, the GPU expansion for “Emmy”
  - “CARO” for the German Aerospace Center (DLR)
  - Max Planck Society (MPG) HPC systems (housing)
  - upcoming: SCC CPU+GPU expansion 2023
- IT staff+systems from GWDG, Uni Göttingen, UMG, MPG
- Expansions in construction, e.g. future CS lecture hall

GPU System “Grete”

New partition for the NHR system “Emmy”

NHR-NORD@GÖTTINGEN

- Installed at RZGö (MEGWARE, 2022)
- Connection to “Emmy” storage
  - Local flash storage (Atos/DDN) for the GPU cluster at RZGö→`/scratch`
  - Link to existing `WORK` (at MDC)→`/scratch-emmy`
  - LNet routers for connecting to Lustre (IB↔eth↔OPA)
- New login node `glogin9`
  - `glogin-gpu.hlrn.de`

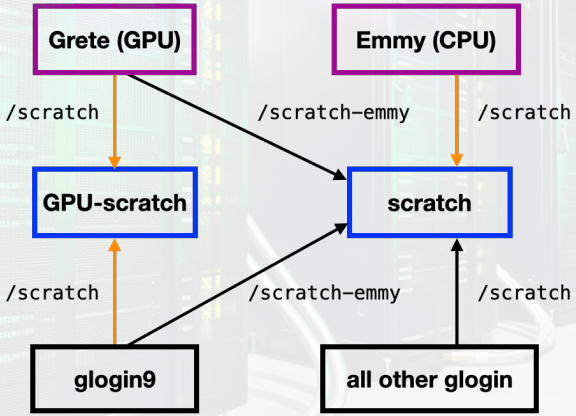
GPU System “Grete”

New partition for the NHR system “Emmy”

- Base system: 34 nodes (3 racks), each equipped 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)
- GPU extension 2023: 5 Knoten, each equipped with
  - 2x Intel Xeon Platinum 8468 (48 “Sapphire Rapids” cores)
  - 512 GB Speicher (DDR4, 3200 MHz)
  - 2x 1,92 TB NVMe SSD
  - 4x NVIDIA H100 GPU (SXM5, 94 GB HBM2 Speicher)
  - 2x Mellanox InfiniBand HCA (HDR)

GPU System “Grete”

Frontend Nodes



How to access fast storage

frontend

storage

compute nodes

access

fast access

GPU System “Grete”
Specifications of the GPUs

NHR-NORD@GÖTTINGEN

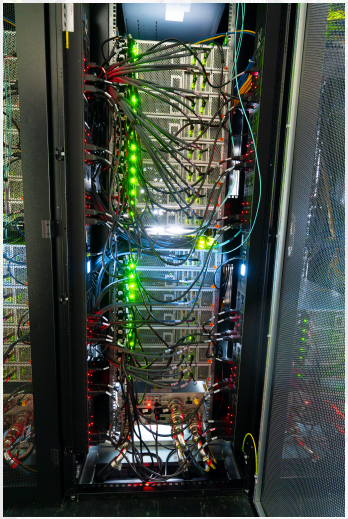
- 136x A100 Tensor Core GPU (Ampere architecture)
  - GA100 GPU: 108 SMX → 6912 CUDA cores, 432 Tensor cores
  - 19,49/9,746/155,92 TFLOPs (SP/DP/Tensor)
  - 40 GB VRAM (HBM2e)
  - SXM4 form factor, 4 GPUs on “Redstone” baseboard
  - NVLINK full mesh, 800 GB/s bidirectional bandwidth per GPU
- GPUS can be requested individually in the same batch system (Slurm)

“Grete”/REACT
back

Source: GWDG

“Grete”/REACT
front

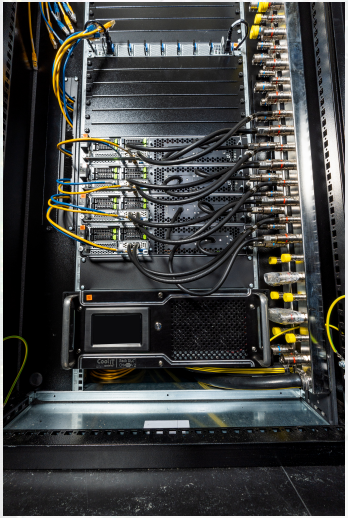
Source: GWDG



“Grete”/REACT

power+fabric

Source: GWDG

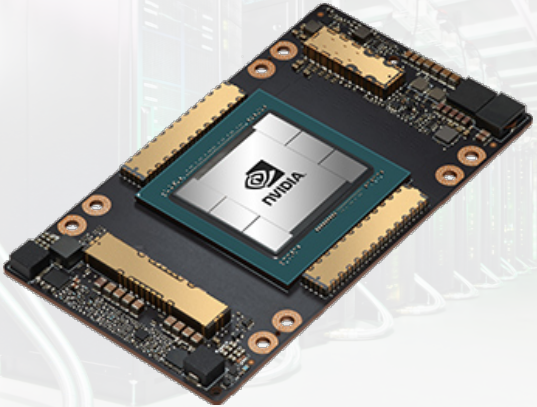


“Grete”/REACT

LAN+DLC

Source: GWDG

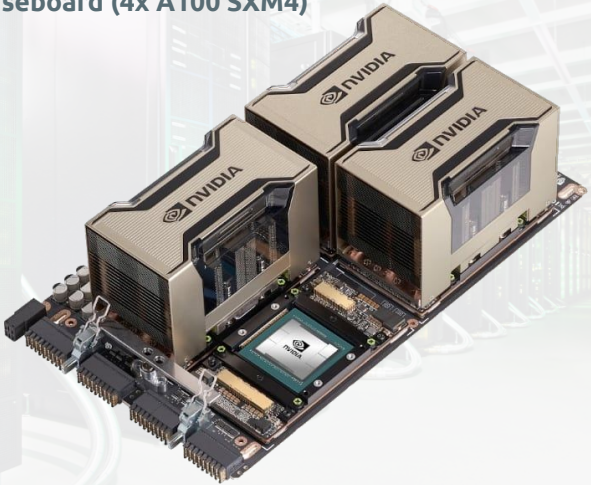
GPU example
NVIDIA A100 SXM4



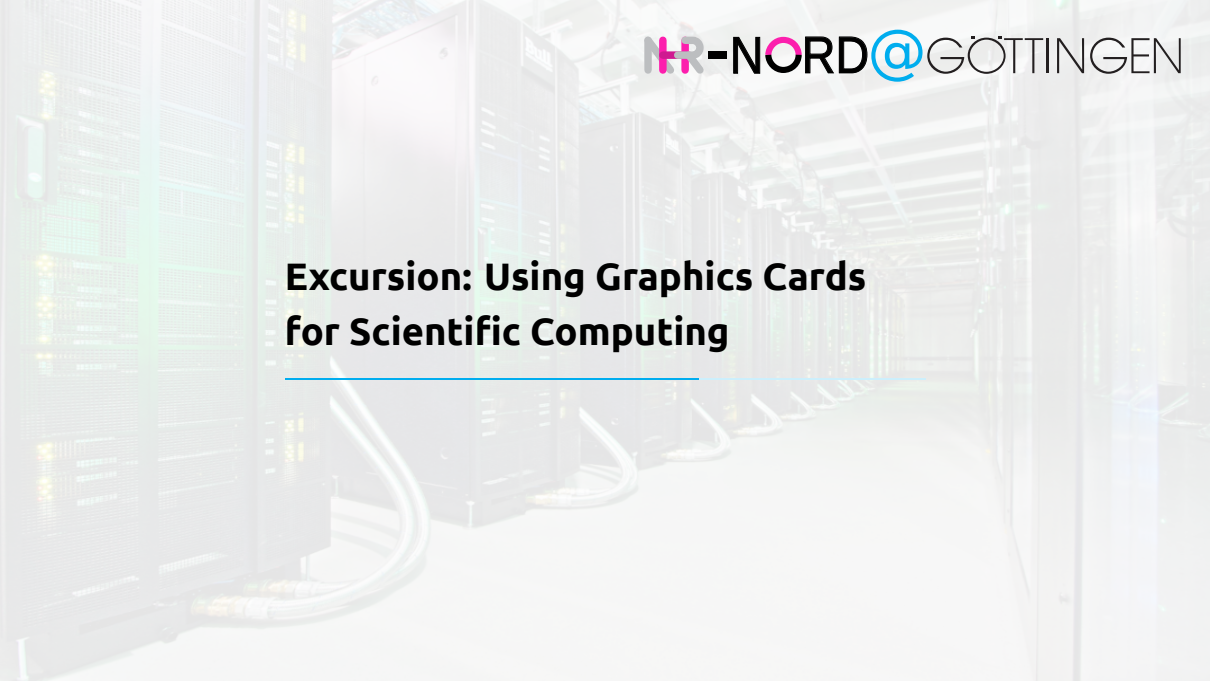
Source: NVIDIA

GPU example

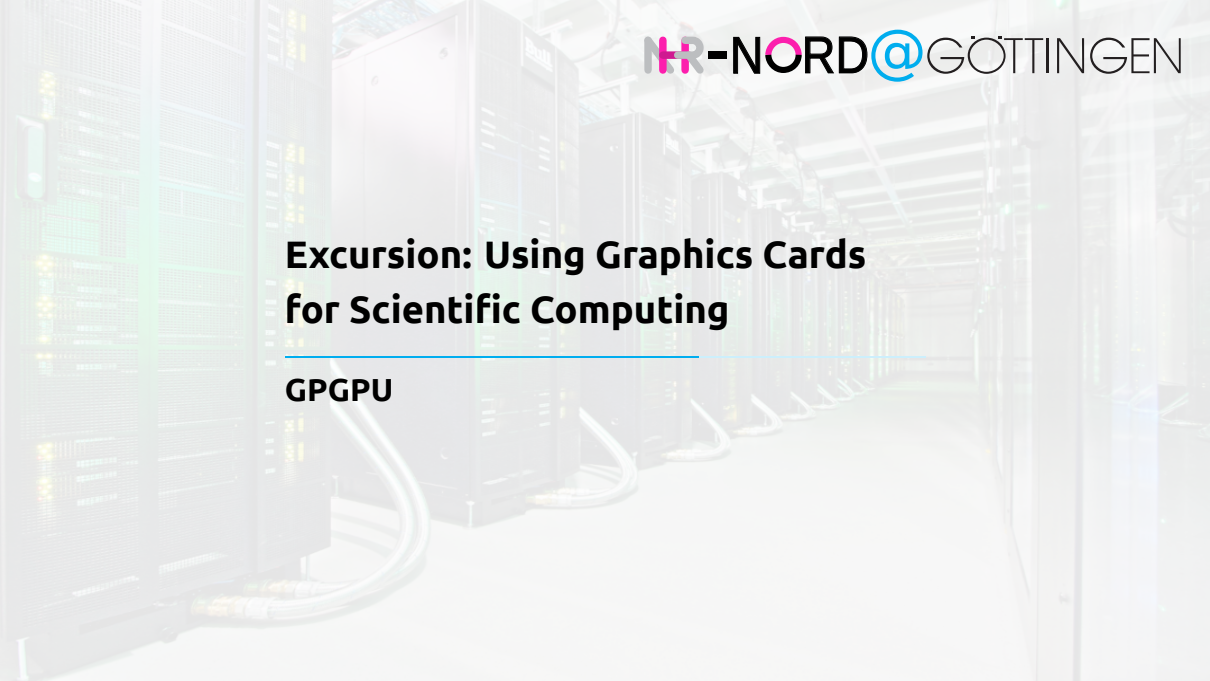
NVIDIA “Redstone” Baseboard (4x A100 SXM4)



Source: NVIDIA



Excursion: Using Graphics Cards for Scientific Computing



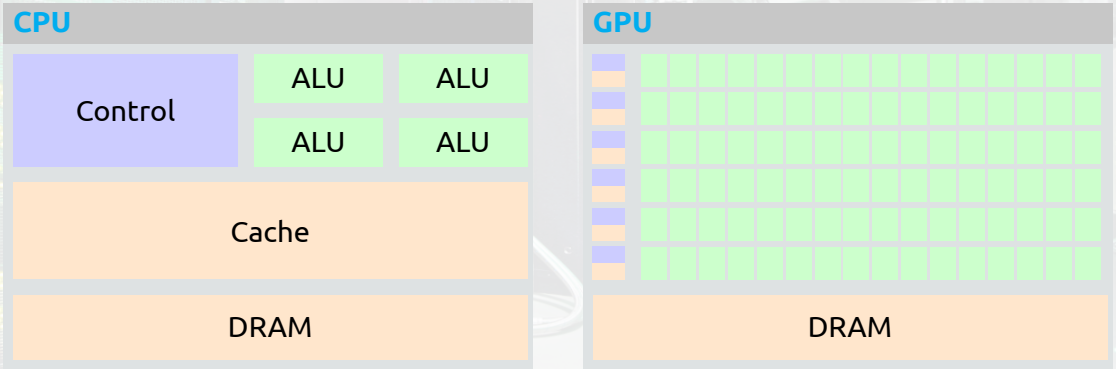
Excursion: Using Graphics Cards for Scientific Computing

GPGPU

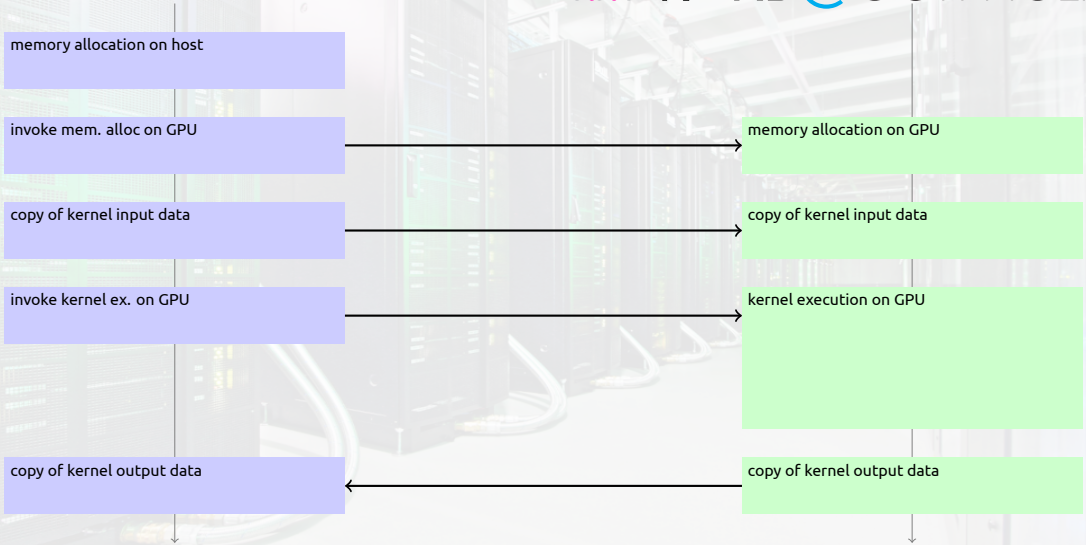
GPUS for Scientific Computing: GPGPU

- Speedup of single-thread execution stagnating (cf. Moore's Law)
- Parallelize software at the thread level
- General Purpose Computation on Graphics Processing Unit
- Graphics Processor is used as Coprocessor

Comparison: CPU/GPU



GPGPU Programming Model



NHR-NORD@GÖTTINGEN

Excursion: Using Graphics Cards for Scientific Computing

Usage

Typical Usage Scenarios

NHR-NORD@GÖTTINGEN

AI Research

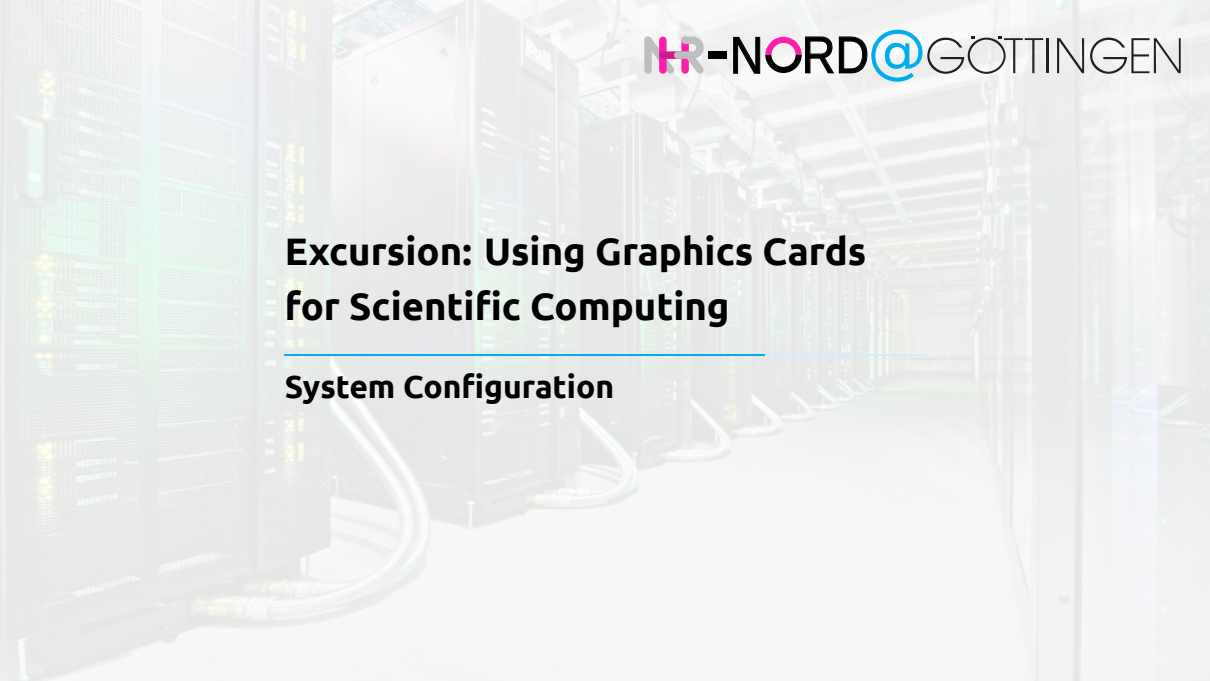
- Labs with many students doing small-scale experiments
- Mid-scale development/experiments by researchers or students
- Large-scale/scalable execution of developed pipelines

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



Excursion: Using Graphics Cards for Scientific Computing

System Configuration

GPU System “Grete”

HLRN-NORD@GÖTTINGEN

- **Partitions** (`-p`, `--partition=<partition_names>`)

Partition	Purpose	Nodes	GPU
`grete`	huge jobs	52	4x A100
`grete:shared`	small jobs	62	4x-8x V100/A100
`grete:interactive`	testing/hyperparameters	3	V100/A100 MIG sl.
`grete:preemptible`	interrupt+resume	3	V100/A100 MIG sl.
`gpu`	small jobs	3	4x V100

- **GPUs** (`-G`, `--gpus=[type:]<number>`)

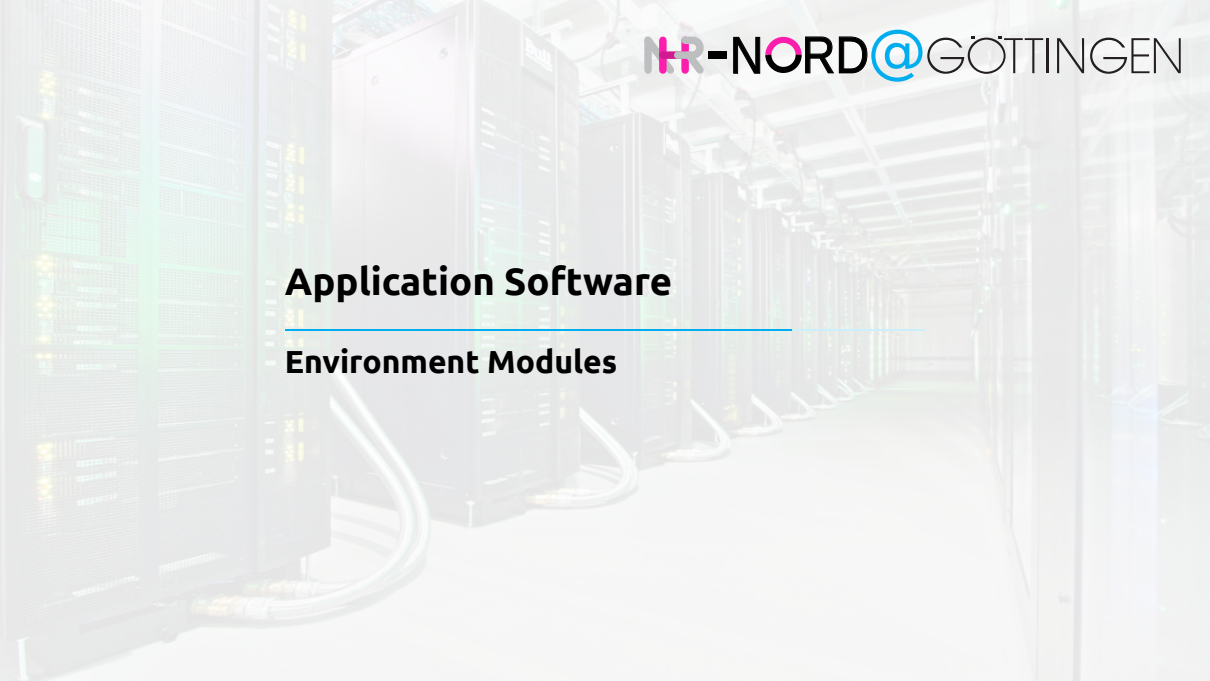
GPU/slice	Per node	Description
`A100`	max. num.	full GPU
2g.10gb	8	2x Compute Instance, 10 GB memory (2 per GPU)
3g.20gb	4	3x Compute Instance, 20 GB memory (1 per GPU)

- Full docs: `https://www.hlrn.de/doc/display/PUB/GPU+Usage`

GPU System “Grete”
Environment Modules

- NVIDIA HPC SDK 23.3: `nvhpc/23.3`
  - for own MPI implementation: `nvhpc-nompi/23.3`
  - ...and also own compiler: `nvhpc-byo-compilers/23.3`
- OpenMPI with CUDA and InfiniBand support
  - HPC-X: `nvhpc-hpcx/23.3`
  - NVIDIA/Mellanox OFED stack: `openmpi-mofed/4.1.5a1`

Application Software



Application Software

Environment Modules

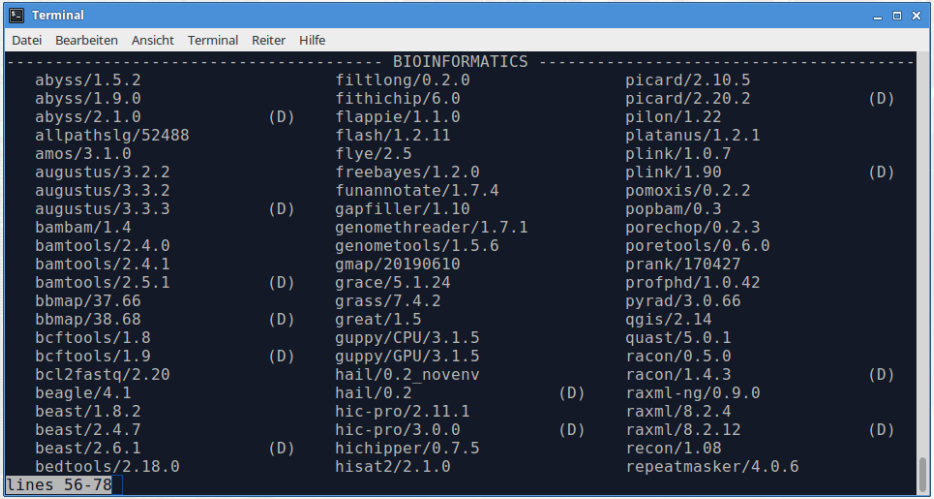
Module system

NHR-NORD@GÖTTINGEN

- 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`

Module system

NHR-NORD@GÖTTINGEN



NHR-NORD@GÖTTINGEN

Application Software

Data Analytics Applications

Data Analytics Applications

NHR-NORD@GÖTTINGEN

- Interactive **Jupyter** notebooks can use HPC as a backend
- **TensorFlow** is a popular AI tool that can use CPU and GPU dynamically
- **Apache Spark** clusters can be spawned as multi-node HPC jobs

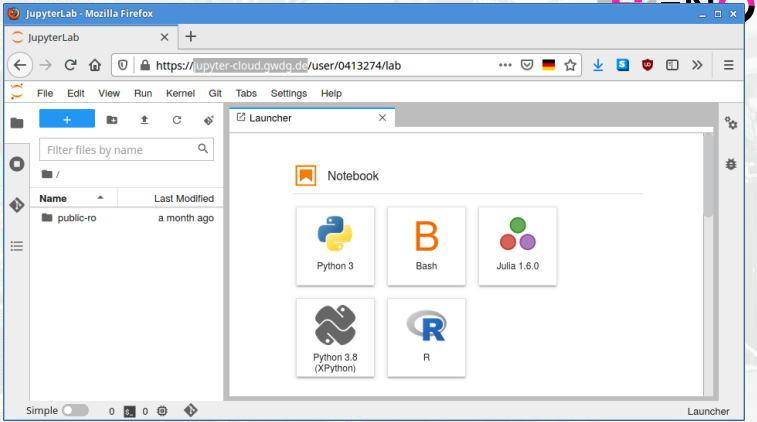
Data Analytics Applications
Apache Spark / Tensorflow

NHR-NORD@GÖTTINGEN

- 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



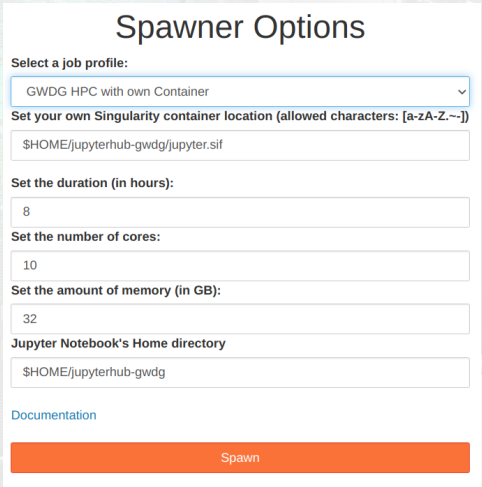
Data Analytics

Jupyter

- JupyerHub on HPC: `https://docs.hpc.gwdg.de/getting_started/jupyterhub`
- JupyterCloud instance: `https://jupyter-cloud.gwdg.de`

Data Analytics Applications
Jupyter on SCC

NHR-NORD@GÖTTINGEN



- `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

Getting access

NHR-NORD@GÖTTINGEN

Getting access

Account creation

Usage Scenarios Tier 2/3

NHR-NORD@GÖTTINGEN

- **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 (User/affiliation 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

NHR-NORD@GÖTTINGEN

- **Scientific Compute Cluster (SCC)**
  - Account activation: request via hpc@gwdg.de
  - Job prioritization via FairShare
- **NHR systems “Emmy”/“Grete”**
  - Apply for an account at nhr-support@gwdg.de
  - Starting from 300.000 Coreh per quarter
    - Project application for further compute time
    - NHR-wide portal JARDS: https://jards.nhr-verein.de/
    - 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)

Further resources

Where to get more information?

NHR-NORD@GÖTTINGEN

- **GWDG HPC docs** `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
- General GWDG support address: support@gwdg.de
- Rocket.Chat channel `#hpc-users` (`https://chat.gwdg.de`)

Courses

NHR-NORD@GÖTTINGEN

- We offer **courses** on SCC usage, programming with MPI, OpenMP, CUDA, Python, Research Data Management and more
  → **GWDG Academy** `https://www.gwdg.de/academy`
- Regular introductory course **“Using the Scientific Compute Cluster”**
  - will be held online via BigBlueButton
- Yearly event: **“Parallel Programming Day”**
  `https://user.nhr.zib.de/wiki/spaces/PUB/pages/429900/Parallel+Programming+Day+series`

Courses

NHR-NORD@GÖTTINGEN

- We offer **courses** on SCC usage, programming with MPI, OpenMP, CUDA, Python, Research Data Management and more
  → **GWDG Academy** `https://www.gwdg.de/academy`
- Regular introductory course **“Using the Scientific Compute Cluster”**
  - will be held online via BigBlueButton
- Yearly event: **“Parallel Programming Day”**
  `https://user.nhr.zib.de/wiki/spaces/PUB/pages/429900/Parallel+Programming+Day+series`
- **Thanks for your attention! Questions?**