HLRN/BremHLR User Forum

July 10, 2023
15 s.t.

Agenda

1. BremHLR - the Bremen Supercomputing Competence Center and its activities
2. NHR - National High Performance Computing Association
3. The new application scheme JARDS
4. GPU-Systems at HLRN
5. Questions, feedback, discussions

BremHLR

Bremen Supercomputing Competence Center

BremHLR

- Founded when HLRN-I was operational (Juli 2003)
- Bases on cooperation contracts
- Cooperating partners:
  - Alfred-Wegener-Institute Bremerhaven (Computing Center, Branch Computing)
  - University Bremen (ZARM, ZeTeM)
  - Constructor University Bremen (CLAMV)
  - University of Applied Sciences Bremen
  - University of Applied Sciences Bremerhaven
- Support by Senatorin für Wissenschaft und Häfen
  - Office and lead consultant
- Location of Office: University of Bremen, ZeTeM
- General goal:
  - Support of HPC projects in Bundesland Bremen as part of the competence network of HLRN

BremHLR-Consultants

- Thorsten Coordes (Engineering)
- Paul Gierz (Climate system modeling, high-level analysis languages)
- Achim Gelessus (Chemistry, Local consultant Constructor University)
- Lars Nerger (BremHLR-Coordinator, Earth System Sciences, Physics)

(consulting is generally part-time)

⇒ First point of contact in case of problems & questions

⇒ You can also write to support@hlrn.de

⇒ If you can ‘t reach „your“ consultant:
contact coordinator (bremhlr@uni-bremen.de)

BremHLR-Activities

- Consulting, e.g.
  - checking of computing time proposals
  - help in case of compute problems
  - help in case of allocation problems
- Events
  - User forum
  - Course “Parallel Programming with MPI and OpenMP”
    - October 9-13, 2023 @ University of Bremen
      - registration open

NHR

„Nationales Hochleistungsrechnen e.V.“
„National High Performance Computing Association“

NHR - https://www.nhr-verein.de

- A new organization for Tier-2 computing centers
  - NHR Alliance founded in 2021
  - 9 NHR computing centers (based on application phase)
    - different topical foci
  - Changed funding scheme
  - Currently still in restructuring phase
    - E.g. HLRN Scientific Council stepped down; re-established with new members
    - Establishing cooperations,
      - e.g. Atomistic Simulation Center was founded (virtual center: Paderborn, Erlangen, and Berlin)
      - preparations for Earth System Simulation Center

NHR Centers

- IT Center, RWTH Aachen
- Hochschulrechenzentrum, Technische Universität Darmstadt
- Zentrum für Nationales Hochleistungsrechnen Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg
- GWDG, University Göttingen (*HLRN*)
- Steinbuch Centre for Computing (SCC), Karlsruher Institut für Technologie
- Zentrum für Informationsdienste und Hochleistungsrechnen, Technische Universität Dresden
- Zuse Institute Berlin, Berlin University Alliance (*HLRN*)
- Paderborn Center for Parallel Computing, University Paderborn
- NHR Süd-West, Goethe-Universität Frankfurt, Technische Universität Kaiserslautern, Johannes Gutenberg-Universität Mainz, Universität des Saarlandes

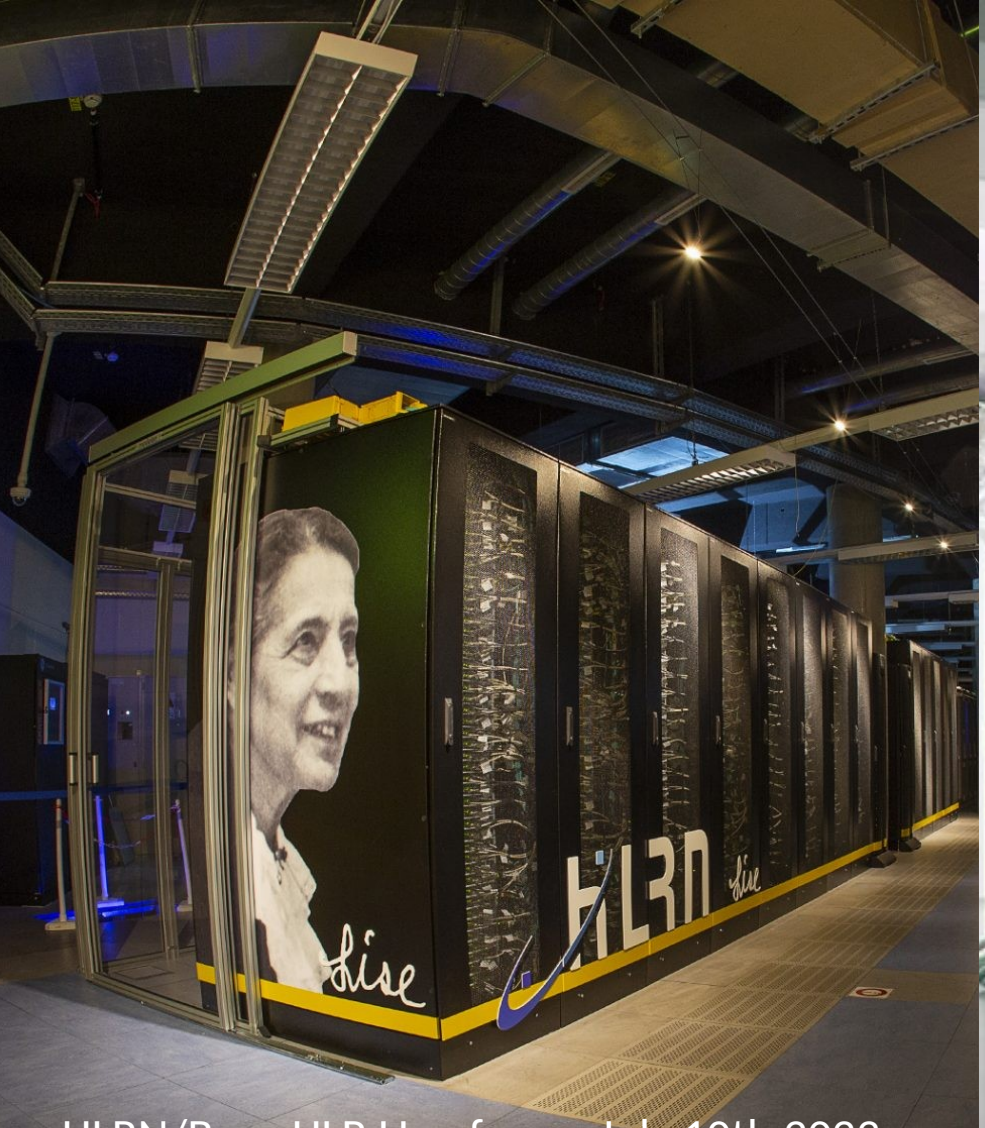
NHR - https://www.nhr-verein.de

- Implications for users
  - Application possible at any of the computing centers
  - Harmonized application deadlines
  - Unified application system (JARDS, more later)
  - NHR-wide training opportunities (subscribe to NHR mailing list)

  - HLRN (North German Supercomputing Alliance) expected to continue (Even though both computing centers are now more independent)
  - BremHLR will get broader scope to support use of more NHR centers

The new application scheme JARDS

Stefan Wollny, HLRN/NHR Berlin



HLRN/BremHLR Userforum July 10th 2023
Project applications

Stefan Wollny

NHR : Association („Verein“)

- Structure :
  - Board of Directors (chairmen)
  - NHR-Office (support)
  - Strategy Committee (→ GWK)
  - “NHR-Betreiberausschuss“ (operators)
  - “NHR-Nutzungsausschuss“ (applications)
  - NHR-Scientifc Advisory Board (internat.)

NHR :

- “NHR-Nutzungsausschuss“

  \

  - Local (each NHR):
    - HLRN: Scientific Council („Wissenschaftlicher Ausschuss“)
    - Has one member in the „Nutzungsausschus“

Therefore …

- Local or central descisions about policies
- Local or central „panels“ (formulars) in JARDS
- Applicants do not need to know this …

NHR :

- One System (JARDS) for all applications
- Apply from now on for computing
  time starting at october 1st
- Deadline July 15th („usually“ : July 1st)

- NHR@ZIB AND NHR@Göttingen
  → Scientific Council („WA des HLRN“)
- NHR-“Normal“ : < 20 Mio coreh/year → local review
- NHR-“Large“ : >= 20 Mio coreh/year → local review + discussion in „Nutzungsausschuss“
- NHR-“Test“ : not needed / user account has 75 kcoreh (up to 300)

NHR :

- Overview:
  - https://www.nhr-verein.de/
- Application:
  - https://jards.nhr-verein.de/
- Daily Usage (e.g. new project members)
  - https://zulassung.hlrn.de/
- Help & Documentation (HLRN)
  - https://www.hlrn.de/doc/display/PUB/Apply+for+a+Compute+Project

NHR :

- Application:
  - https://jards.nhr-verein.de/ - > Live Demo

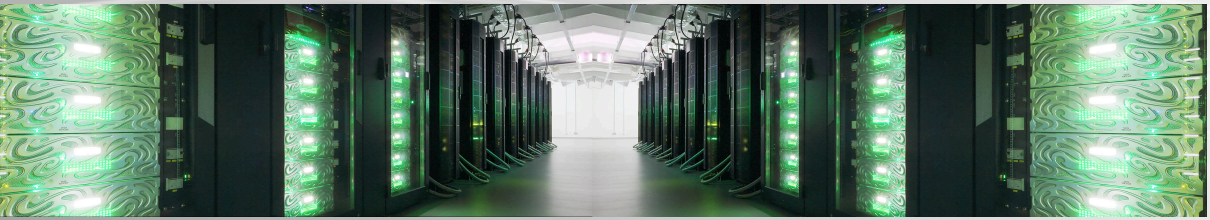
GPU-Systems at HLRN

Timon Vogt, HLRN/NHR Göttingen

GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

New GPUs on NHR - GRETE and GPU A100

AG-C



July 10, 2023

Overview

At the end of 2022 and beginning of 2023,
the NHR systems "EMMY" and "LISE" got new GPU clusters

EMMY’s new GPU cluster: GRETE

EMMY/GRETE

- At the end of 2022, the GWDG got a new GPU cluster delivered
- 34 new GPU based nodes, each with 4 Nvidia A100
- Nodes are named: `ggpu[101-134]`
- This new cluster has been named GRETE
  - After Grete Hermann, Emmy Noether's first doctoral student
- After extensive tests, the systems are now publicly available since beginning of April

Nodes

34 GPU-Nodes, each with:

- 2x AMD EPYC "Milan" 7513 CPU (32 Cores, 2.6 GHz Base clock, 200 Watt TDP)
- 512 GB RAM (16x 32 GB DDR4 3200 MHz ECC-registered Modules)
- 2x 1TB NVMe PCIe 3.0 SSD for temporary files / caches
- 2x ConnectX-6 HDR100 Infiniband Ports
- 4x NVIDIA A100 GPUs, 40GB HBM2 memory, NVLink, 400 Watt TDP
  - connected via NVIDIA Redstone GPU Baseboard

One new, dedicated Login-Node: `glogin9`

- AMD EPYC "Milan" 7313P (16 Cores, 3.0 GHz Base clock, 155 Watt TDP)
- 128 GB RAM (8x 16 GB DDR4 3200 MHz ECC-registered Modules)

NVIDIA A100 SXM4

"The NVIDIA A100 Tensor Core GPU delivers unprecedented acceleration—at every scale—to power the world's highest- performing elastic data centers for AI, data analytics, and high- performance computing (HPC) applications."

https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet-us-nvidia-1758950-r4-web.pdf

NVIDIA A100 TENSOR CORE GPU SPECIFICATIONS (SXM4 AND PCIE FORM FACTORS)

	A100 40GB PCIe	A100 80GB PCIe	A100 40GB SXM	A100 80GB SXM
FP64	9.7 TFLOPS			
FP64 Tensor Core	19.5 TFLOPS			
FP32	19.5 TFLOPS			
Tensor Float 32 (TF32)	156 TFLOPS | 312 TFLOPS*			
BFLOAT16 Tensor Core	312 TFLOPS | 624 TFLOPS*			
FP16 Tensor Core	312 TFLOPS | 624 TFLOPS*			
INT8 Tensor Core	624 TOPS | 1248 TOPS*			
GPU Memory	40GB HBM2	80GB HBM2e	40GB HBM2	80GB HBM2e
GPU Memory Bandwidth	1,555GB/s	1,935GB/s	1,555GB/s	2,039GB/s
Max Thermal Design Power (TDP)	250W	300W	400W	400W
Multi-Instance GPU	Up to 7 MIGs @ 5GB	Up to 7 MIGs @ 10GB	Up to 7 MIGs @ 5GB	Up to 7 MIGs @ 10GB
Form Factor	PCIe		SXM	
Interconnect	NVIDIA® NVLink® Bridge for 2 GPUs: 600GB/s ** PCIe Gen4: 64GB/s		NVLink: 600GB/s PCIe Gen4: 64GB/s	
Server Options	Partner and NVIDIA-Certified Systems™ with 1-8 GPUs		NVIDIA HGX™ A100-Partner and NVIDIA-Certified Systems with 4,8, or 16 GPUs NVIDIA DGX™ A100 with 8 GPUs	

* With sparsity

** SXM4 GPUs via HGX A100 server boards; PCIe GPUs via NVLink Bridge for up to two GPUs

Scratch

- GRETE is located at the RZGoe
  (as opposed to EMMY in the MDC)
- GRETE also has its own scratch, located at the RZGoe (`/scratch-grete`)
  - Built from two DDN ES400NVX with a combined capacity of 130 TiB
  - Connected to GRETE via InfiniBand-Switches with 800 GBit/s
- The EMMY scratch is also available, mounted under `/scratch-emmy`
- The connection to EMMY is made via 4 100GBit/s fibers between the RZGoe and the MDC
  - Two Fibers are for IP Traffic (`/home`, `/sw`, ...)
  - Two Fibers are for Lustre Traffic (`/scratch-emmy`)
- On `glogin9`, the GRETE scratch is also mounted

Software stack

- Operating system: Rocky Linux 8.6
  - Kernel version: 4.18.0
- CUDA (version 11.8) is preinstalled in the images
  - Further CUDA versions are available via the module system
- Further software will be available via the module system
  - Then, specifically build for the GRETE nodes

Slurm integration

- Due to the newer operating system, the GRETE nodes have their own Slurm binaries
- Completely integrated into the existing Slurm cluster
- The nodes will be available in the partition `grete`, `grete:shared` and `grete:interactive`

MIG - Multi-Instance-GPU

With their A100, A30 and H100 GPUs, NVIDIA offers a new functionality: Multi-Instance-GPUs (MIG)

- Allows to completely split a GPU in up to 7 instances
  - $\Rightarrow$ dedicated memory, caches, streaming-processors for each
- For `grete:interactive`, we are using this feature in the `2g.10gb` configuration
  - As for testing and development, a full A100 GPU might not always be necessary
  - Especially during model development in AI, a smaller slice of a GPU usually suffices

LISE’s new GPU partition: GPU A100

LISE/GPU A100

- 42 new GPU based nodes, each with 4 Nvidia A100
- Nodes are named: `bgn[1001-1042]`
- The partition has been named "GPU A100"

Nodes

42 GPU-Nodes, each with:

- 2x Intel Xeon "Ice Lake" Platinum 8360Y (36 cores per socket, 2.4 GHz base clock, 250 W TDP)
- 1 TB RAM (DDR4 3200 MHz ECC-registered Modules)
- 7.68TB NVMe SSD for temporary files / caches
- 200 GBit/s InfiniBand Adapter (Mellanox MT28908)
- 4x NVIDIA A100 GPUs, 80GB HBM2 memory, SXM

Two new, dedicated Login-Nodes: `bgnlogin[1-2]`

- 512 GB RAM (3200 MHz ECC-registered Modules)
- No GPUs (and no CUDA drivers)

Software stack

- Operating system: Rocky Linux 8.6
  - Kernel version: 4.18
- CUDA and further specifically compiled software is available via a special software module
  - `sw.a100`

Slurm integration

- The nodes will be available in the partition `gpu-a100`, `gpu-a100:shared` and `gpu-a100:shared:mig`

MIG - Multi-Instance-GPU

With their A100, A30 and H100 GPUs, NVIDIA offers a new functionality: Multi-Instance-GPUs (MIG)

- Allows to completely split a GPU in up to 7 instances
  - $\Rightarrow$ dedicated memory, caches, streaming-processors for each
- The MIG slices are available on the `gpu-a100:shared:mig` partition, with the `1g.10gb` configuration

Further questions, feedback, discussions...

Further Questions?

bremhlr@uni-bremen.de

See also:
www.bremhlr.uni-bremen.de