GWDG
Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen

Efficient Data Center Operation

Modeling of a Warm Water Cooling System

Sebastian Krey

28.07.2025

Outline

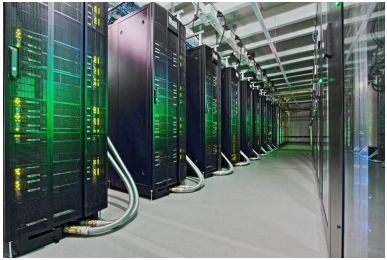
About GWDG

- IT service center and data center operation for **University Göttingen** and **Max Planck Society** (MPG) since 1970
- Operating site of “North German Supercomputing Alliance” (**HLRN**) since 2018, since 2021 part of **NHR**
- AI Service Center **KISSKI** for critical infrastructure
- HPC operating site for the “German Aerospace Center” (**DLR**) since 2022

Network for National High-Performance Computing

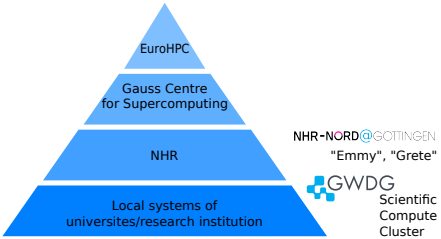
NHR-NORD@GÖTTINGEN

UGOE + GWDG



- Since 2021: Funding for national Tier-2 supercomputing (62.5M=C p.a.)
  - Nine centres
  - Annual funding 7,3M€ p.a.
- Usable for researchers at all German universities
  - Up to 1,200k CPU core/1500 GPU hours p.a. usable without application
  - Secure Workflow for processing sensitive data (medical, financial, etc.)
  - Larger projects require application
    → `https://docs.hpc.gwdg.de/start_here/nhr_application_process/index.html`

HPC systems at GWDG



- Tier 2: **HLRN/NHR “Emmy”**
  Top500 #47 Nov. 2020, now #219
- Tier 2: **NHR/KISSKI “Grete”**
  Top500 #141 Nov. 2023, Green500 #22, now #226/#70
- Tier 2: **NHR/KISSKI “Grete Phase 3”**
  Top500 #274 Nov. 2024, Green500 #24, now #315/#35
- Tier 3: **Scientific Compute Cluster**
- **“CARO”** for **DLR**
  Top500 #135 Nov. 2021, now #335
- Several smaller systems for MPG and UGOE

HLRN-IV “Emmy”

Uni Göttingen/GWDG

- TOP 500: #47 in 2020-11 (5.95 PFlop/s), now #226,
  approx (inofficial) 4.56 GFlops/Watt (would have been #55 in 2020-11)
- phase 1 compute nodes (air cooled), out of operation
  - 2x Intel Xeon Gold 6148 (SKL), 40 cores per node, 480 GB SSD
  - 432x 192 GB, 16x 768 GB
  - 240 kW
- phase 2 compute nodes (warm water DLC), EoL H1 2026
  - Intel Walker Pass System
  - 2x Intel Xeon Platinum 9242 (CLX-AP), 96 cores per node
  - 1100x 384 GB, 16x 768 GB, 2x 1536 GB
  - 80-85% CoolIT DLC
  - 1100 kW

NHR “Emmy Phase 3”

- Replacement of Emmy Phase 1
- 447 nodes
- 2 Sapphire Rapids 48 core CPUs (Xeon Platinum 8468)
- Memory: 164x256GB, 32x1TB, 3x2TB, remaining 512GB
- Cornelis Omnipath 100G interconnect
- Connection to storage of other islands via routing
- 65-75% CoolIT DLC with direct free cooling with outside air for residual heat
- 450 kW

NHR “Grete+”

- GPU cluster consisting of three procurement modules
- Performance optimized: 5.46 PFlop/s
- Energy optimized: 34.647 GFlop/Watt (best in Germany at inauguration)
- 103 nodes
- 2 AMD Epyc Milan 7513
- 4 A100 GPUs per node (36 nodes with 40 GB, 2 nodes 8xA100)
- Dual rail Infiniband HDR interconnect
- Cluster local GPU Direct enabled storage
- 70% CoolIT DLC
- 205 kW performance optimized, 128 kW energy optimized

NHR “Grete Phase 3”

- Performance optimized: 3.65 PFlop/s
- Energy optimized: 53.708 GFlop/Watt
- 25 nodes
- 2 Intel Sapphire Rapids 8468
- 4 H100 GPUs per node
- Dual rail Infiniband HDR interconnect
- Cluster local GPU Direct enabled storage
- 70% CoolIT DLC
- 80 kW performance optimized, 58 kW energy optimized

DLR “CARO”

- Operated for the German Aerospace Center
- 1370 nodes with 2 AMD Epyc Rome 7702
- 3.46 PFlop/s, TOP 500 #135 in 2021-11, now #228
- 364 TB memory
- 24 Quadro RTX 5000 for visualization
- Infiniband HDR100 interconnect
- 8.4 PiB DDN Lustre (200 TiB SSDs)
- 55% CoolIT DLC, high temperature air cooling for residual heat
- 760 kW

Storage Systems

- WORK MDC: 7 Celestica SC6100 1.7 PiB NVME
- WORK RZGÖ: DDN ExaScaler 6 510 TiB NVME 2x ES400NVX
- HOME/SW/WORK KISSKI: VAST Data 1.1PiB NVME (3x dBox, 3x cBox)
- WORK SCC: 2.2 PiB BeeGFS based on DDN SFA7990 block storage
- WORK Ceph: 600 TiB NVME CephFS and S3
- COLD: 20 PiB HDD CephFS and S3
- HSM/Tape: Quantum StorNext HSM 60+ PiB

Data Center Efficiency

- until early 2000s very little relevance
- Power Usage Effectiveness (PUE) as first KPI established in 2007
- Back then PUE >2 quite normal
- Steady decrease until about 2019
- Currently stagnation around 1.5x (worlwide average)
- Direct liquid cooling allows PUE below 1.1 (practical limits at about 1.03)
- Data center cooling systems are complex
- Interactions of components heavily impact total efficiency

Real world problems

- 800 kW HPC system connected to 800 kW warm water cooling circuit
- HPC and cooling system declared for 35°C water inlet
- After first full year of operation insane amount of water consumption
- Way too many hours with adiabatic cooling
- Discussion with cooling system operator:
  - Cooling system unstable
  - Reduction of inlet temperatur to 25°C necessary for stable operation

Real world problems

- Every component has certain performance characteristics
- Small changes in one or multiple parameters can have dramatic effects
- Data sheets usually state performance on one or two operating points
- Different example operating points for different components
- Result of 250-510 kW cooling capacity for a system consisting of:
  - Two 500 kW coolers
  - One 800 kW plate heat exchanger
  - One 5000 l buffer tank
  - Two 750 kW plate heat exchangers for 760 kW HPC compute
  - Several pumps inbetween the components

Why so complicated?

Simple basic formulas for heat transport:

$$Q = m \cdot c_p \cdot \Delta t$$

with Q heat flow in watts, $m = V\rho$ the mass flow, c_p the specific heat cpacity and the temperature difference Δt

and heat.transer via a heat exchanger:

$$Q = k \cdot A \cdot \Delta t_{m,log}$$

with the heat exchange coefficient k, the transfer area A the the average logarithmic heat difference $\Delta t_{m,log}$.

Interactions

Output of every component is input of another component.

$\rightarrow$ Changing one parameter requires recalculation of everything.

Example from above:

	cooler	plate heat exchanger		HPC system
inlet $°C$	50	34	50	35
outlet $°C$	35	49	35	48
flow m^3/h	2 x 28,7 = 57,4	47	47	37
capacity kW	2 x 500 = 1000		800	760

Manual recalculation with real operating parameters:

510 kW instead of assumed 800 kW cooling capacity

Buffer tank

In theory buffer tank has no impact on the cooling capacity.

Serial connection has no effect on capacity, but requires matching flow rates on both sides of the buffer.

Parallel connection does not require matching flow rates $\rightarrow$ easier operation, often chosen. But:

If flow rates do not match $\rightarrow$ turbulences in the buffer $\rightarrow$ mixture of the water $\rightarrow$ reduction of Δt $\rightarrow$ strong effect on cooling capacity

250 kW cooling capacity without matching of flow rates

Buffer tank model

Very little information available $\rightarrow$ simple linear model

Input variable: Absolute value of difference between flowrates $|V_l - V_r|$

Target value: Temperature increase/reduction $\Delta t_{bb}/\Delta t_{bt}$

$$\rightarrow T_{rl} = T_{rr} - \Delta t_{bt} \text{ and } T_{ir} = T_{il} + \Delta t_{bb}$$

Shiny App

Small Demo

Summary

- Efficient data center operation is important but complex
- Every component has influence on total performance
- Read the details in the specifications
- Changing one parameter can change the whole system (even with hydraulic separation)
- Manual recalculation tedious work
- Small app can simplify the work of the data center operations team