Karlsruher Institut für Technologie

2nd HPC Café

07.02.2025

Agenda HPC Café – 07.02.2025

1. Follow up on the upcoming HPC system „HoreKa-2“
   - Status Procurement
   - Results of the survey on scientific needs for „HoreKa-2“

2. News about NHR@KIT
   - Project self service
   - Update on granting information about compute projects
   - AI fast track

3. Questions and Answers

Karlsruher Institut für Technologie

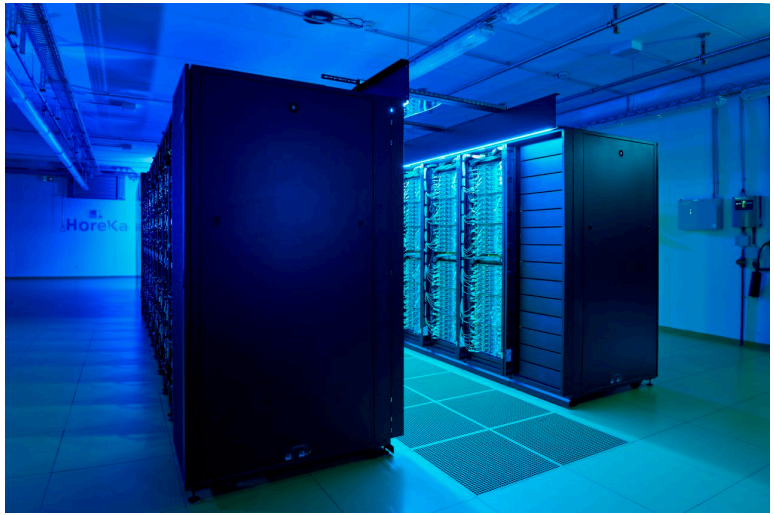
SCC

1.a. Follow up on

Robert Barthel (SCC)

HoreKa-2 Overview

- **Budget**
  - ~15 million €
  - ~**2.4** million € HAICORE 3.0
- **Procurement:** March 2025
- Commissioning of first phase: End of Q4/2025
- **Full commissioning: 2026**
- **Components:** Compute + Filesystem
- **Location:** Campus North
  - DLC, hot water cooled, ~40°C in, ~45°C out
  - Power envelope: less then 1 MW



Procurement Considerations

Basic considerations:

- Tier-2 system
- As technologically open as possible
- Has to serve both HPC and AI workloads
- Energy Awareness ⟶ Active Powercapping

Procurement Considerations Implications

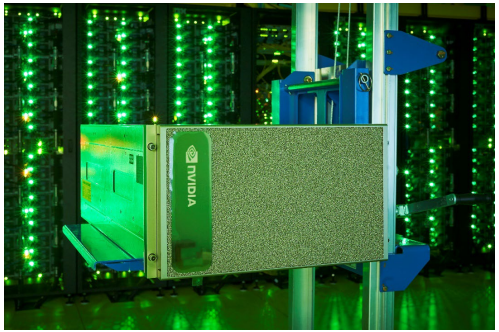
Tier-2:

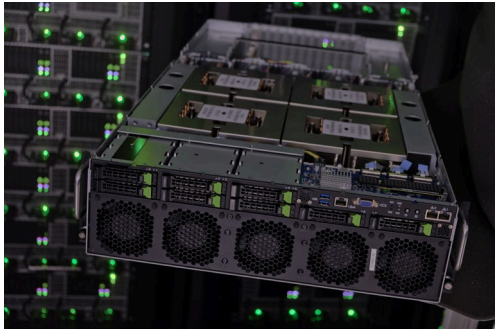
- Cluster-Size
- ~~Exotic hardware conceivable~~ ←→ Tier-3
- Advanced users assumed

Procurement Considarations Implications

Has to serve both HPC and AI workloads:

- ~~If~~ **hybrid** system ⟶ yes
  - **Split** between CPU and accelerated partitions ⟶ defined
  - Same host architecture on CPU/GPU-nodes?
    ⟶ most likely not
    - x86/x86?
    - x86/ARM?
    - ARM/ARM?
- Accelerators
  - **Double precision** required? / „HPC-Flavor“
    (yes: rather AMD)
  - **Convenience**/off-the-shelf software? / „AI-Flavor“
    (yes: rather NVIDIA)

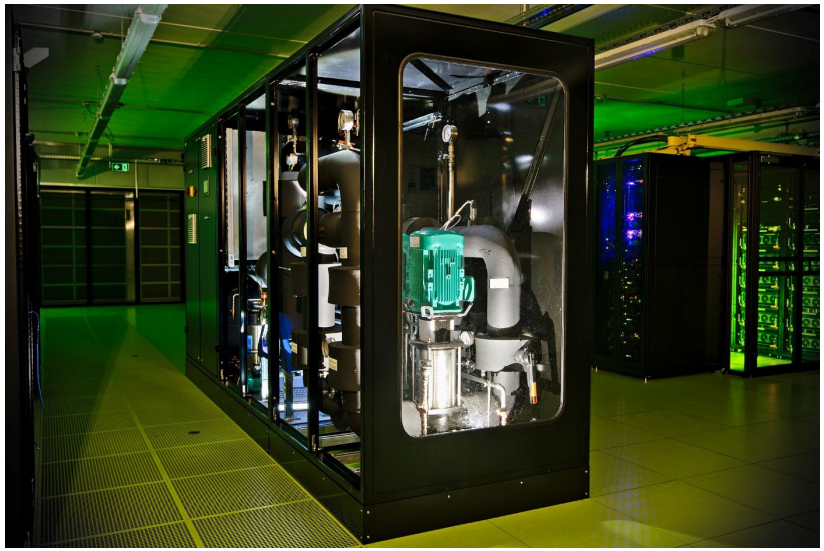




Procurement Considerations Implications

Energy Awareness:

- Fixed **financial budget** for energy
  - Checkpointing
  - Power scaling ⟶ most likely enabled
  - CPU/GPU-hours → **energy budget** for compute projects
- Full DLC components preferred
- Accelerated codes!
- Please contact SSPE-Team ;)



Procurement Considerations: Update

- Result of Market analysis (nothing surprising…)
  - **FLOP/Watt** → in favor of **GPU**, ~ factor 5
  - **FLOP/€** → in favor of **GPU**, ~ factor 2
  - **∑Energy** → in favor of **GPU**, CPU-only system would exceed energy envelope

  → We would get a GPU-only system

- Pragmatic approach
  - “Freezing” of the total performance of the CPU partition, ~ 2.5 PFLOP
  - „Backfill“ with GPU nodes

Karlsruher Institut für Technologie

SCC

1.b. Results of PI survey on scientific needs of HoreKa-2

Funda Elewa (SCC)

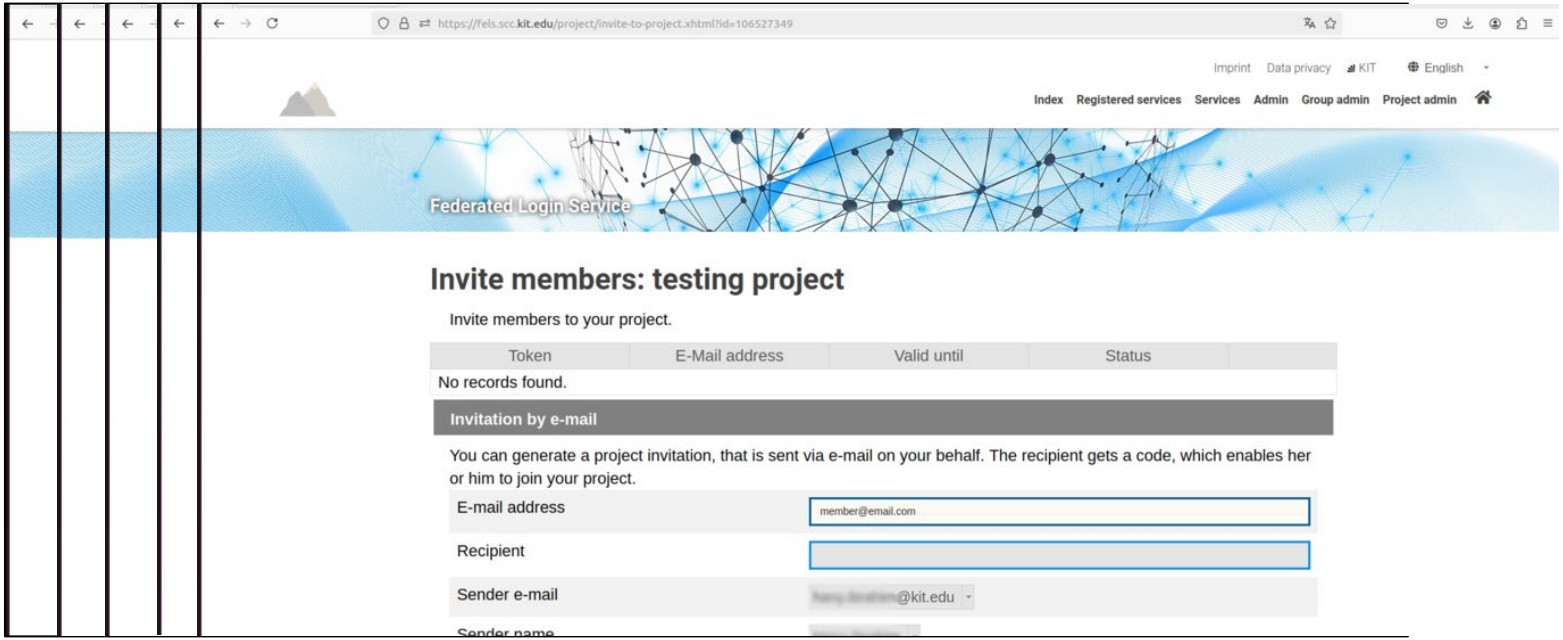
Karlsruher Institut für Technologie

SCC

2. News about NHR@KIT

2.a. Project self service

- https://www.nhr.kit.edu/604.php



2.b. Update on granting information about compute projects

- Taking into account the current invest & operating costs (especially electricity), one CPU core hour and GPU hour on the HoreKa system corresponds to a monetary equivalent of 0.026 and 0.032 euros, respectively. Your overall project therefore has a monetary equivalent of ***<value> Euro***.

- Energy-efficient HPC is one of the goals of NHR@KIT. One CPU Core hour and GPU hour corresponds to an energy of about 0.014 and 0.052 kilowatt-hours (kWh), respectively. Using all granted computing time of your project therefore corresponds to a total energy of ***<value> kWh***. According to the conversion factors https://www.umweltbundesamt.de/publikationen/green-cloud-computing, this corresponds to ***<value> tons of Carbondioxide***. The NHR@KIT offers various support services to increase efficiency. If required, please contact: hpc-dic-support@scc.kit.edu.

2.c. AI fast track

- New project category:

Project category	Duration	Effectively Granted Resources	Review	Call
NHR Test	6 months, not extendable	500.000 CPUh / 5.000 GPUh	Technical	Rolling
NHR Starter	1 year, not extendable	360.000 CPUh / 10.000 GPUh	Technical	Rolling
AI Fast Track	1 year, extendable	380.000 CPUh / 20.000 GPUh	None	Rolling

Karlsruher Institut für Technologie

SCC

3. Questions and Answers