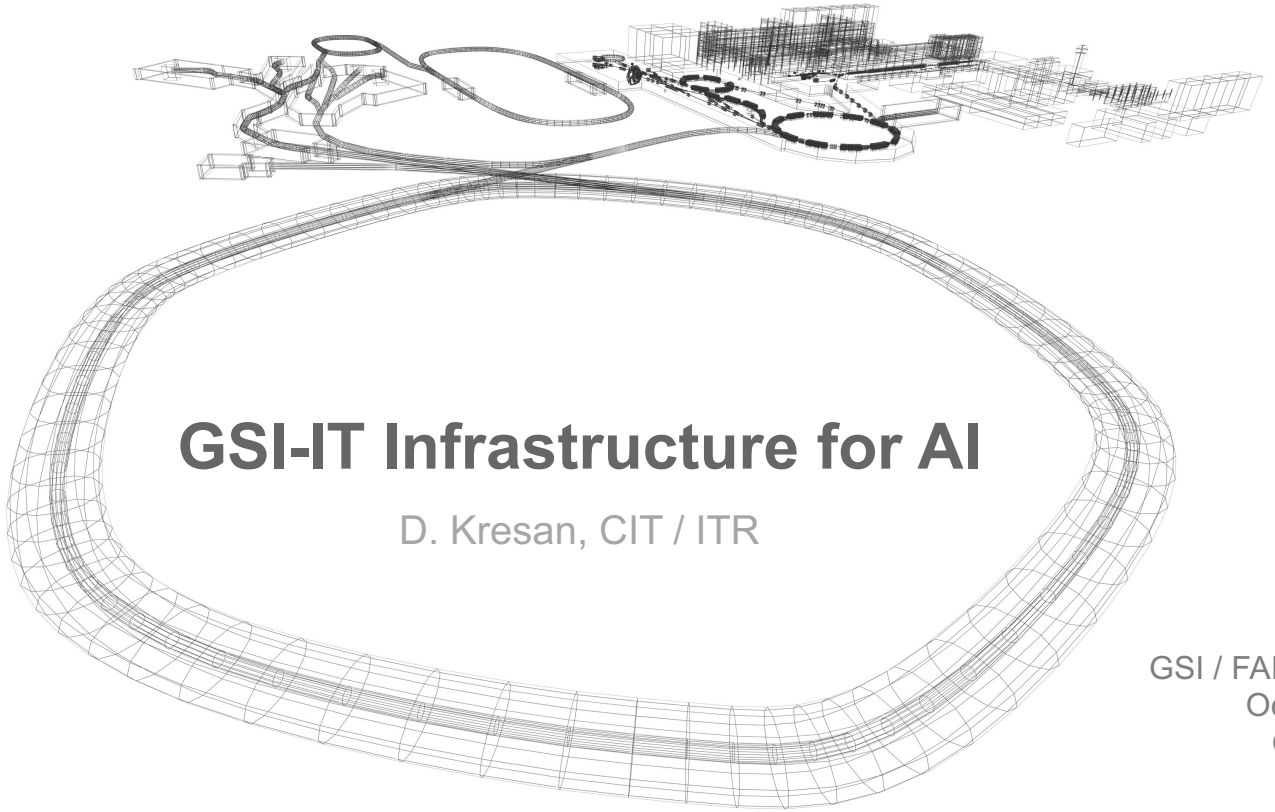
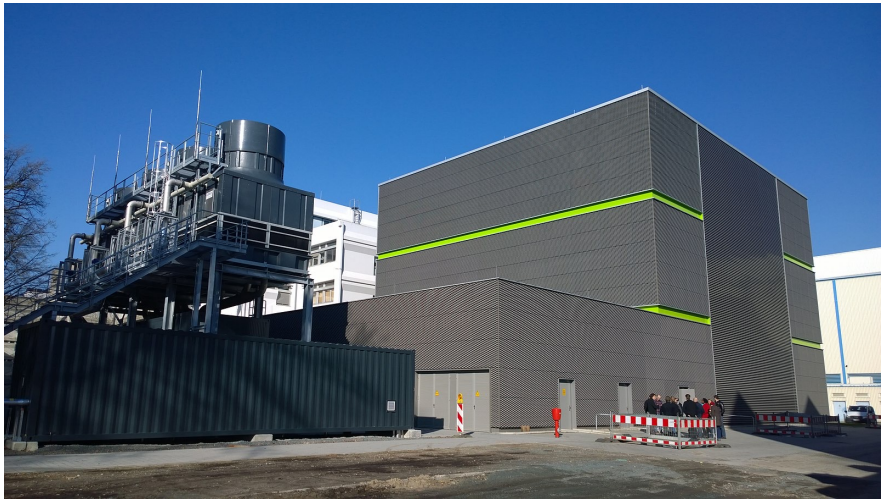


GSI-IT Infrastructure for AI

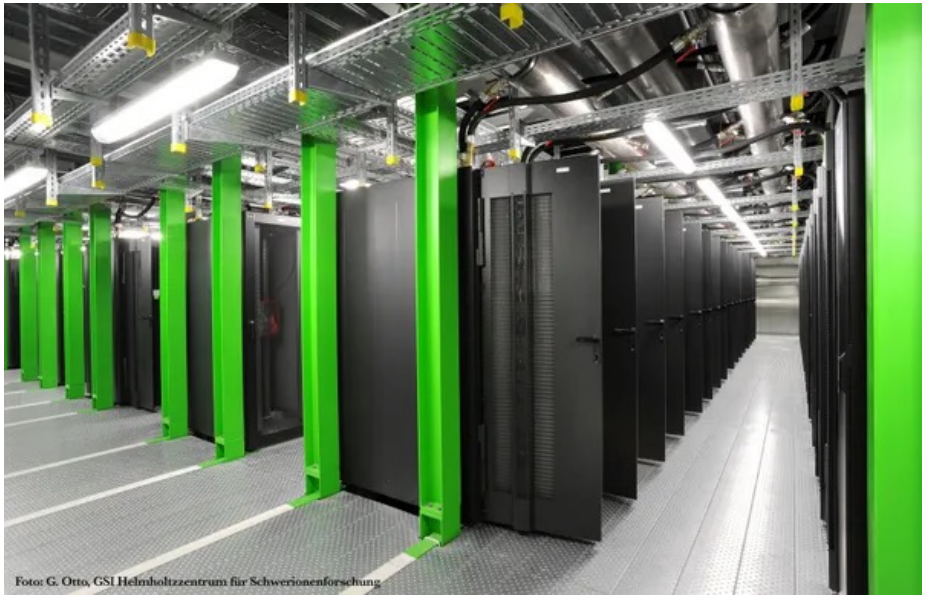
D. Kresan, CIT / ITR

GSI / FAIR AI Workshop
October 29, 2024
GSI, Darmstadt

Green IT Cube



In operation since 2016



PUE < 1,07
4 MW cooling
Capacity for 768 racks on 6 floors

Computing cluster “Virgo”

- 495 nodes
  - 233 Intel Xeon Gold, 48 cores, RAM 4 GB/core
  - 45 AMD EPYC 7662, 128 cores, RAM 8 GB/core
  - 135 AMD EPYC 7713, 128 cores, RAM 4 GB/core
  - 82 AMD EPYC 9654, 192 cores, RAM 4 GB/core
- 50 k phys. cores or 100 k CPUs (with hyperthreading)
- 200 Gb/s HDR InfiniBand internal network

Computing cluster “Virgo”

- 495 nodes
  - 233 Intel Xeon Gold, 48 cores, RAM 4 GB/core
  - 45 AMD EPYC 7662, 128 cores, RAM 8 GB/core
  - 135 AMD EPYC 7713, 128 cores, RAM 4 GB/core
  - 82 AMD EPYC 9654, 192 cores, RAM 4 GB/core

Heterogeneous hardware
- Scales up the booting / installing infrastructure
- Provide hardware optimized binaries

- 50 k phys. cores or 100 k CPUs (with hyperthreading)

- 200 Gb/s HDR InfiniBand internal network

Computing cluster “Virgo”

- 495 nodes
  - 233 Intel Xeon Gold, 48 cores, RAM 4 GB/core
  - 45 AMD EPYC 7662, 128 cores, RAM 8 GB/core
  - 135 AMD EPYC 7713, 128 cores, RAM 4 GB/core
  - 82 AMD EPYC 9654, 192 cores, RAM 4 GB/core

Heterogeneous hardware
- Scales up the booting / installing infrastructure
- Provide hardware optimized binaries

- 50 k phys. cores or 100 k CPUs (with hyperthreading)

- 200 Gb/s HDR InfiniBand internal network only, no Ethernet

Distributed storage based on Lustre

- Usable capacity 45 PB
- 37 % utilized

- Distributed over 112 file servers

- 3 metadata servers

- Beginning 2025: additional 150 PB for extension / replacement

Storage systems at GSI

Name	Path	Submit nodes	Worker nodes	Description
Home-directory	/u/$USER	Read / Write	- -	Private directory of a user
CVMFS	/cvmfs	Read *	Read	For software distribution
Scratch	/scratch	Read / Write	Read	For building software
Shared storage	/lustre	Read / Write	Read / Write	For large data files

(*) Publishing on CVMFS requires dedicated host with permissions granted to 1 person / repository / publisher

GPU cluster

- 50 nodes with 8 GPU cards/node
- 400 AMD MI100
- ROCm version 6.x
- 7 days max run-time for jobs
- Exclusive GPU allocation

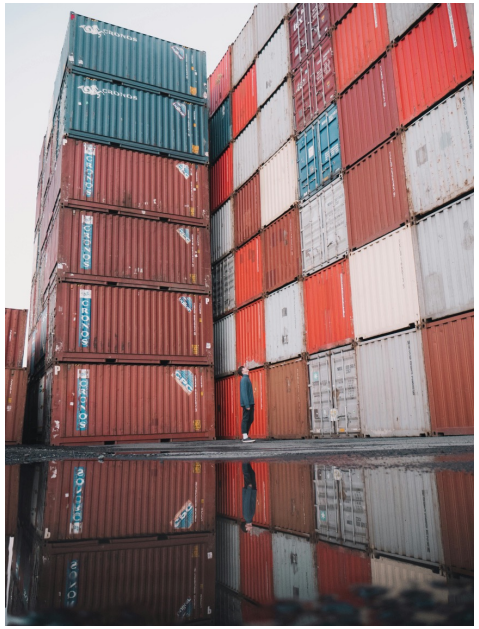
SLURM as resource management system

- We keep cluster updated: SLURM 23-11 is deployed

- 6 partitions for flexible allocation of resources
  - debug: max 30 min
  - main: max 8 hours
  - long: max 7 days
  - high_mem: max 7 days, 8 GB/core RAM
  - gpu: max 7 days, access to GPU

Fully Containerized Approach

- Separate user application space from host system
- Jobs are executed in a container
- **Minimal host system**: HW drivers +
  Slurm and Apptainer



Advantages

- Users are free to choose Linux flavor and install any required software
  - Flexibility in supporting different use cases and workflows
- Admins are free to upgrade host OS and/or Slurm at any time
  - Makes Virgo cluster more scalable

Bare Metal Submitter Node

- `ssh virgo.hpc`
- Submit job in container

Virtual Application Environment (VAE)

- Ready-to-use solution provided by GSI-IT

- `ssh vae24.hpc`
- Login into container - interactive session on submitter node
  - Edit, compile, test, debug

- Submit job which will run in the same VAE

- Fully transparent to a user and easy to use
  - SPANK plugin for Slurm starts container in the background

Scientific software: packages / requests

- More than 700 software packages available in the current version of VAE
- Template form for users to request new software

```
I hereby request the installation of software on the [GSI virgo cluster](https://hpc.gsi.de/virgo/) as described in the
following:

(fields marked with * are mandatory)
* **Name\***: (e.g. ROOT)
* **Version\***: (e.g. 6.20.04)
* **Homepage**: (e.g. https://root.cern)
* **License**: (e.g. LGPL)
* **Known packages**: (e.g. unknown / `root` in Fedora 32 / `root@6.20.04` in Spack)
```

https://git.gsi.de/SDEGroup/SIR

Scientific software: management

- Use Spack to build and install them: https://github.com/spack/spack

- Package manager developed for HPC systems
- Handling of dependencies
- Support of
  - Multi architectures
  - Multi compilers
  - Multi versions
  - Mixed toolchains

- Why all this is good for you?

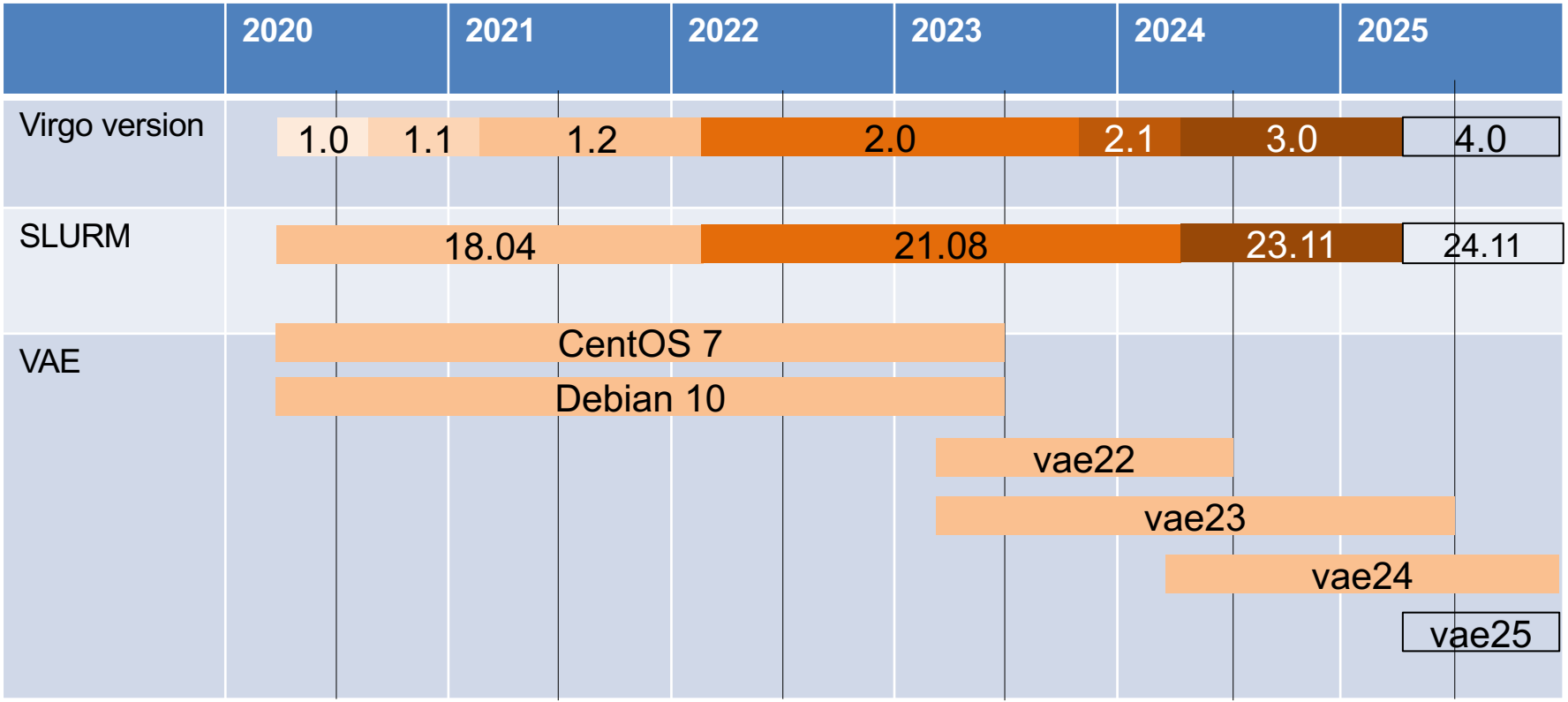
Simplify life of users with custom workflows (i.e. AI workflows)

- Pull container with base OS and drivers from the upstream registry
  - e.g. Docker Hub
- Put on top additional software, tools and models
- Launch a job in your own container on GPU resources

Using CVMFS for software distribution

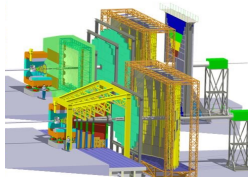
- Build and install software in your CVMFS repository
- Distribute it to the local batch farm or even outside GSI
- Profit from cached access on the client side

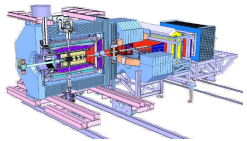
Cluster timeline

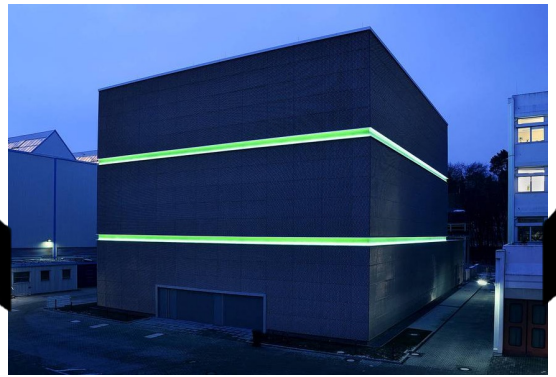


Computing at FAIR: The resources in the Green Cube will be shared between the different FAIR/GSI Partners

Dynamically allocated resources for exclusive usage and limited time







Analysis Facilities

ALICE

No sperate hardware for the online clusters of CBM and PANDA

Generic batch farm for GSI/FAIR Users

CBM

R3B

panda

HADES

AGATA ADVANCED GAMMA TRACKING ARRAY

BIOMAT

APPA

. . . . .

Summary

- 5 years of stable operation with Virgo cluster
- Virgo is highly scalable generic computing cluster which supports modern workflows
- https://hpc.gsi.de/virgo/
- cluster-service@gsi.de