Karlsruhe Institute of Technology

GridKa: LK II, WLCG Tier1, and more

Matter and Universe Days 2022, Darmstadt

Matthias J. Schnepf on behalf of GridKa | 21. October 2022

Supportet Experiments

- ATLAS
- ALICE
- Auger
- BaBar
- Belle
- CMS
- Compass
- LHCb
- IceCube
- XFEL

Facts about GridKa

- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter

Facts about GridKa

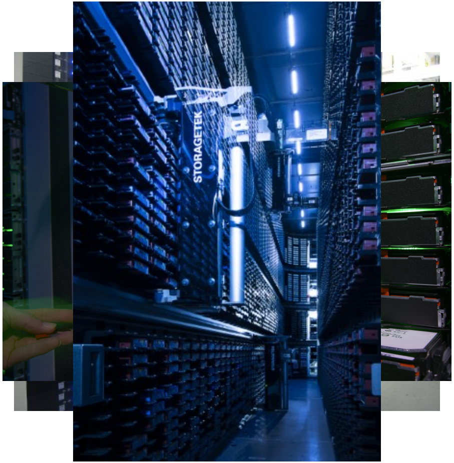
- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter
- computing
  - 450 worker nodes
  - about 59.000 CPU cores

Facts about GridKa

- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter
- computing
  - 450 worker nodes
  - about 59.000 CPU cores
- storage
  - about 48 PB (upgrade to 99 PB) disk storage
  - about 6.600 hard drives

Facts about GridKa

- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter
- computing
  - 450 worker nodes
  - about 59.000 CPU cores
- storage
  - about 48 PB (upgrade to 99 PB) disk storage
  - about 6.600 hard drives
  - about 70 PB tape storage
  - about 10% of the LHC and Belle II data are storage at GridKa



Facts about GridKa

- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter
- computing
  - 450 worker nodes
  - about 59.000 CPU cores
- storage
  - about 48 PB (upgrade to 99 PB) disk storage
  - about 6.600 hard drives
  - about 70 PB tape storage
  - about 10% of the LHC and Belle II data are storage at GridKa
- network
  - 2×100 $\mathrm{Gbit\,s^{-1}}$ WAN
  - 2×100 $\mathrm{Gbit\,s^{-1}}$ direct to CERN

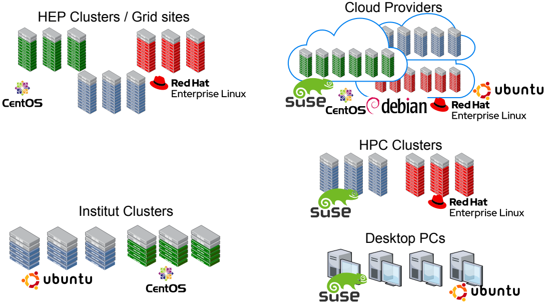
Facts about GridKa

- GridKa is a
  - Helmholtz LK II
  - WLCG Tier1
  - Belle II RAW datacenter
- computing
  - 450 worker nodes
  - about 59.000 CPU cores
- storage
  - about 48 PB (upgrade to 6 age
  - about 6.600 hard
  - about 70 R
  - about 10% HC and Belle II data are storage at GridKa
- network
  - $2\times100\,\mathrm{Gbit\,s^{-1}}$ WAN
  - $2\times100\,\mathrm{Gbit\,s^{-1}}$ direct to CERN

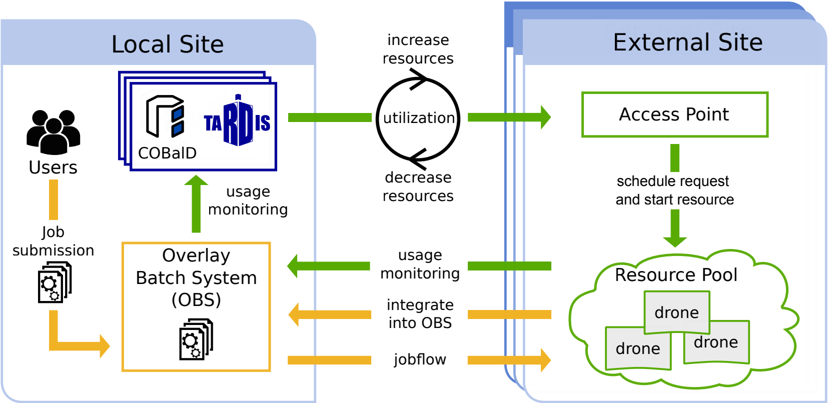
GridKa can and does more

Additional Resources for HEP

- small resource and resources that are not designed for HEP (opportunistic resources) can be used
  - institute cluster
  - cloud provider
  - HPC cluster
  - desktop PCs
- challenges
  - complex resource scheduling due to heterogeneous resource pool
  - software environments provision
  - single point of entry for all resources
  - transparent usage

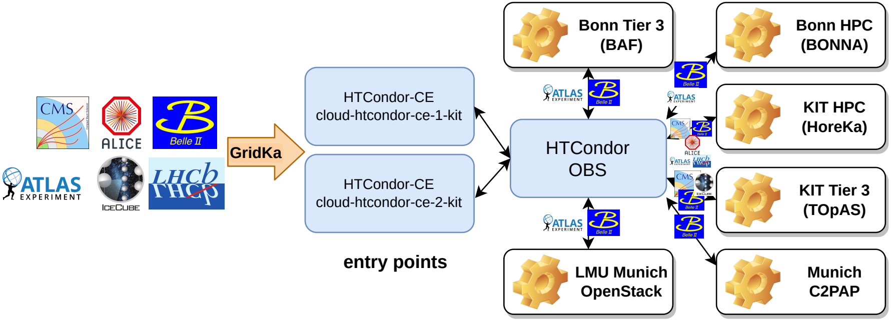


Resource Management: COBalD & TARDIS



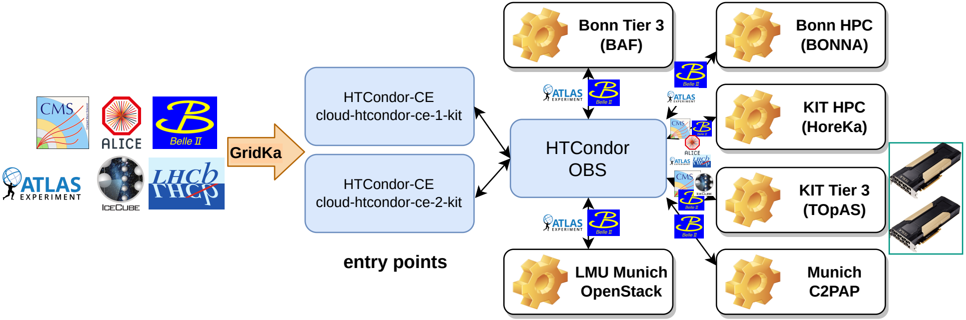
- load balancing daemon COBalD (COBalD - the Opportunistic Balancing Daemon)
- life cycle management TARDIS (Transparent Adaptive Resource Dynamic Integration System)

"Cloud" Computing Resources



- transparent provisioning of computing resources to specific collaborations, see monitoring
- container or virtual machines provide HEP software environment on heterogeneous resources
- integration of further resources in the future - fully transparent and experiment independently

"Cloud" Computing Resources



- transparent provisioning of computing resources to specific collaborations, see monitoring (with GPUs)
- container or virtual machines provide HEP software environment on heterogeneous resources
- integration of further resources in the future - fully transparent and experiment independently

GPUs at GridKa

- more and more applications use GPUs

GPUs at GridKa

- more and more applications use GPUs
- hen egg problem
  - sites do provide resources which are needed
  - experiments develop software for resources that are available

GPUs at GridKa

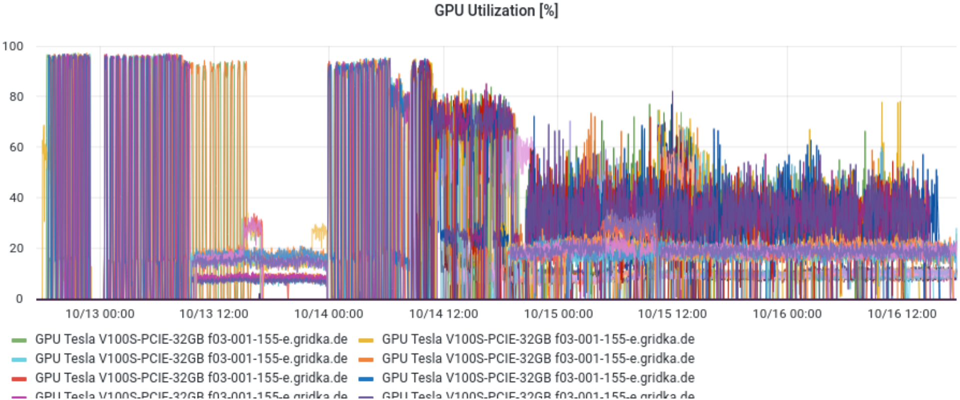
- more and more applications use GPUs
- hen egg problem
  - sites do provide resources which are needed
  - experiments develop software for resources that are available
- end-user analysis cluster with GPUs at GridKa
  - 8x NVIDIA V100
  - 24x NVIDIA V100s
  - 24x NVIDIA A100
- accessible via the physics institute batch system and GridKa cloud CEs

GPUs at GridKa

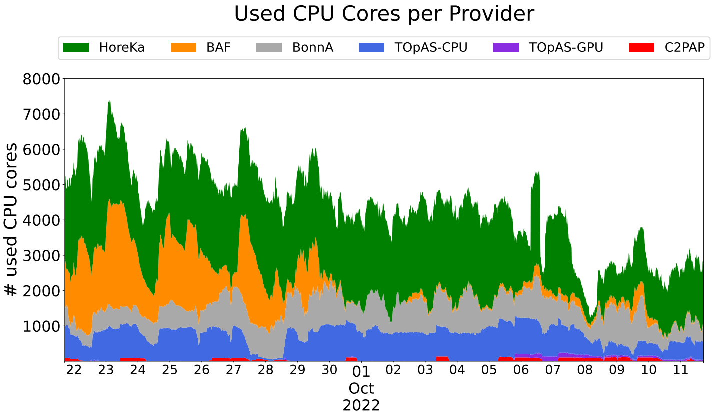
- more and more applications use GPUs
- hen egg problem
  - sites do provide resources which are needed
  - experiments develop software for resources that are available
- end-user analysis cluster with GPUs at GridKa
  - 8x NVIDIA V100
  - 24x NVIDIA V100s
  - 24x NVIDIA A100
- accessible via the physics institute batch system and GridKa cloud CEs
- the local KIT group can use the GPUs and experiments can develop and use the GPUs for/via Grid

GPUs at GridKa: Usage

- used by local CMS and Belle II group as well as CMS via Grid
- ALTAS and Belle II are testing usage via Grid
- development project with local group to increase GPU utilization

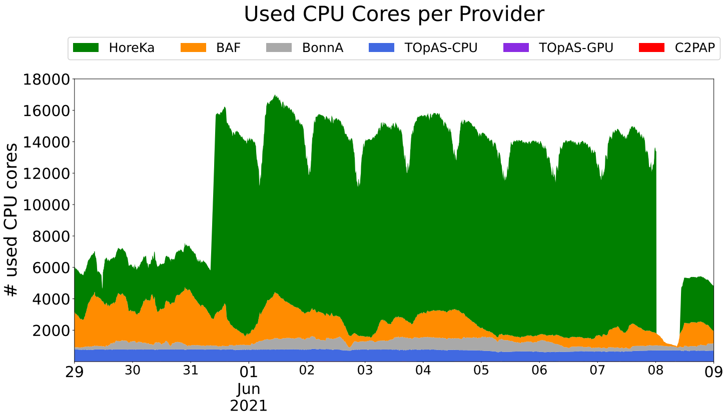


Cloud Resources Provided



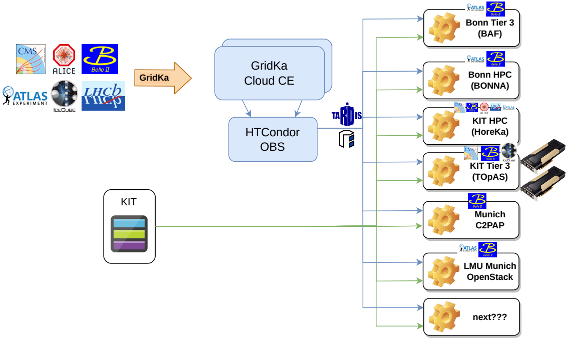
- about 4000 CPU cores additional cores on average

Cloud Resources Provided: Scaletest



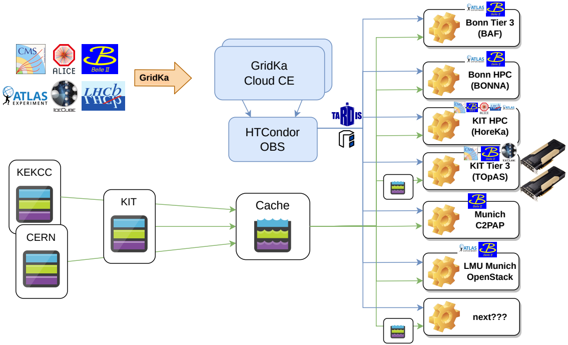
- scaletest with up to **17400** CPU cores

GridKa HEP Cloud



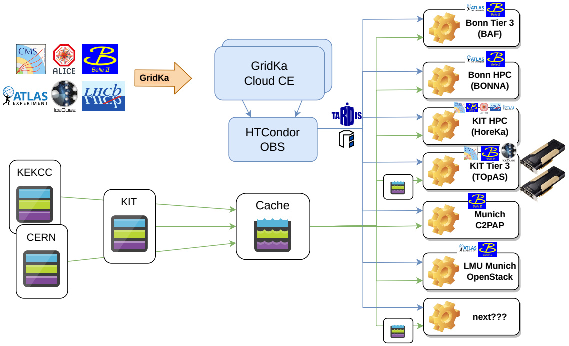
- network connection between GridKa storage and opportunistic computing resources can influence CPU efficiency

GridKa HEP Cloud



- network connection between GridKa storage and opportunistic computing resources can influence CPU efficiency
- caches at the computing resources could help by insufficient network connection

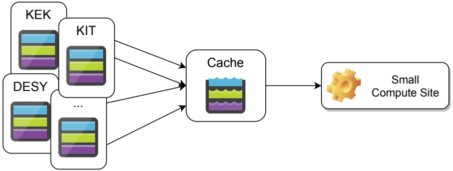
GridKa HEP Cloud



- network connection between GridKa storage and opportunistic computing resources can influence CPU efficiency
- caches at the computing resources could help by insufficient network connection
- development project with the physics institute at KIT to cache Belle II data from other sites

GridKa as Background Storage

- managed Grid storage is expensive and need person power
- small sites can not or would not provide a full Grid storage
- cache with background storage e.g., GridKa could be an alternative for small sites



Conclusion

- GridKa provides a massive amount of computing and storage resources, including GPUs to high energy and astroparticle physics
- development with the physics institute at KIT on caching and resource scheduling
- GridKa provides transparent access to computing resources from partners
- GridKa is ready to be a data hub

Backup

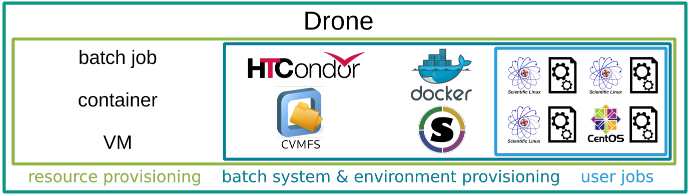
What We Provide

- COBalD & TARDIS
  - https://github.com/MatterMiners/cobald
  - https://github.com/MatterMiners/tardis
- help to setup OBS or integrate site
  - hands on sessions (integration of C2PAP cluster Munich within 4h)
- puppet module
  - https://github.com/unibonn/puppet-cobald
- wlcg-wn container
  - https://hub.docker.com/r/matterminers/wlcg-wn
  - https://github.com/MatterMiners/container-stacks/blob/main/wlcg-wn

```
pip install cobald-tardis
```

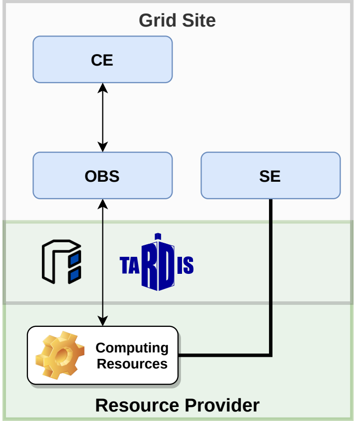
Generalized Pilot Concept

- pilot concept
  - placeholder job allocates resources
  - worker node instance of an Overlay Batch System (OBS) starts payload jobs inside the pilot job
  - requires software environment
- generalized pilot concept ⇒ drone concept
  - resource allocation as
    - batch job
    - virtual machine
    - container
  - provides full Grid software environment
  - drone/pilot/job can run inside a drone

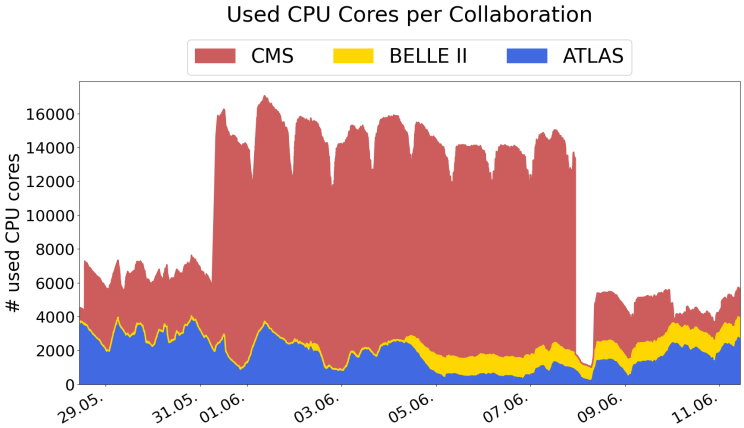


Minimal Setup

- Grid Site
  - standard Grid site services
    - CE
    - OBS for resources
  - provide performant SE and outgoing network
- computing resource provider
  - accessible via HTCondor, Slurm, OpenStack, ...
  - virtualization or container with enables userspace
- COBalD/TARDIS instance
  - lightweight - multiple instances fit on one VM
  - needs just python and resource access
  - instances can be run by Grid site, resource provider, and third party



Provided Resources

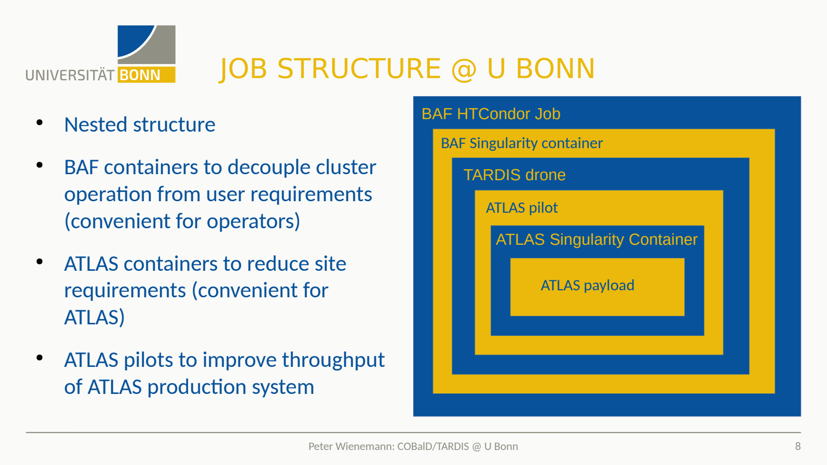


- used by several collaborations
- up to 17.400 CPU cores integrated

Supported Providers

- adapter to interact with provider
- providers
  - HTCondor
  - Moab
  - Slurm
  - CloudStack
  - OpenStack
  - Kubernetes
- further developments are welcome

Pilot inside a Drone



UNIVERSITÄT BONN

JOB STRUCTURE @ U BONN

- Nested structure
- BAF containers to decouple cluster operation from user requirements (convenient for operators)
- ATLAS containers to reduce site requirements (convenient for ATLAS)
- ATLAS pilots to improve throughput of ATLAS production system

BAF HTCondor Job

BAF Singularity container

TARDIS drone

ATLAS pilot

ATLAS Singularity Container

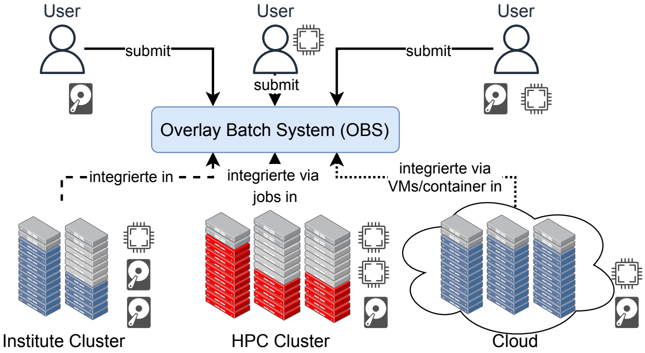
ATLAS payload

Peter Wienemann: COBalD/TARDIS @ U Bonn 8

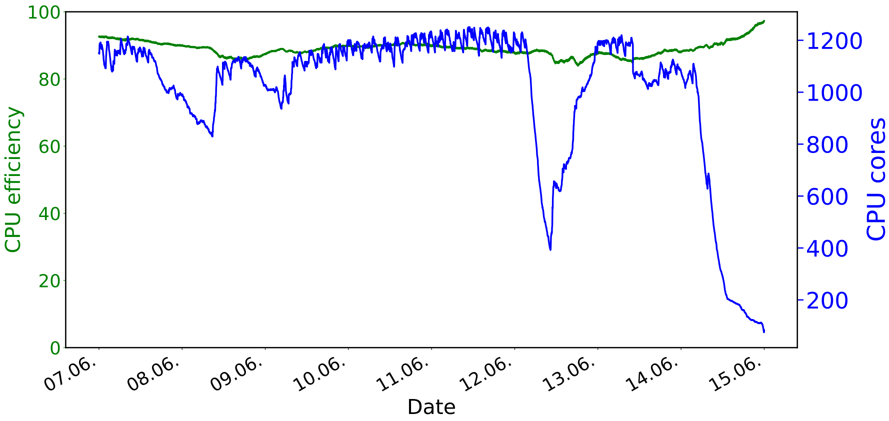
Talk: Opportunistic Resource Mangement with COBalD/TARDIS at U Bonn from Peter Wienemann at the IDT-UM Meeting 30. Sep. 2019: https://indico.physik.uni-muenchen.de/event/22/

Integration of Resources

- integration via drone (virtual machine, container, batch job) into OBS
- HEP software environment provided by virtualization and container technology



Used CPU cores and efficiency for Belle II


CPU efficiency
CPU cores
Date