

# COBaLD/TARDIS - Collaborative and Experiment Overarching Achievements

ErUM Data IDT Collaboration Meeting - 02.04.2020

Manuel Giffels, R. Caspart, M. Fischer, E. Kuehn, M. Schnepf, F. v. Cube, G. Quast

Institute of Particle Physics (ETP) & Steinbuch Centre for Computing (SCC)



# INTRODUCTION

# Opportunistic Resources and their Challenges

## Opportunistic Resource

Any resources **not permanently dedicated** to but **temporarily available** for a specific task, user or group. → No reliance upon WLCG policies!

# Opportunistic Resources and their Challenges

## Opportunistic Resource

Any resources **not permanently dedicated to** but **temporarily available for** a specific task, user or group. → No reliance upon WLCG policies!

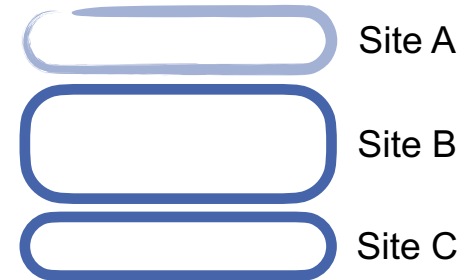
- Each opportunistic resource is different (very heterogenous system)



Where to send my jobs?



Which resources are available?  
Which resources are suitable?

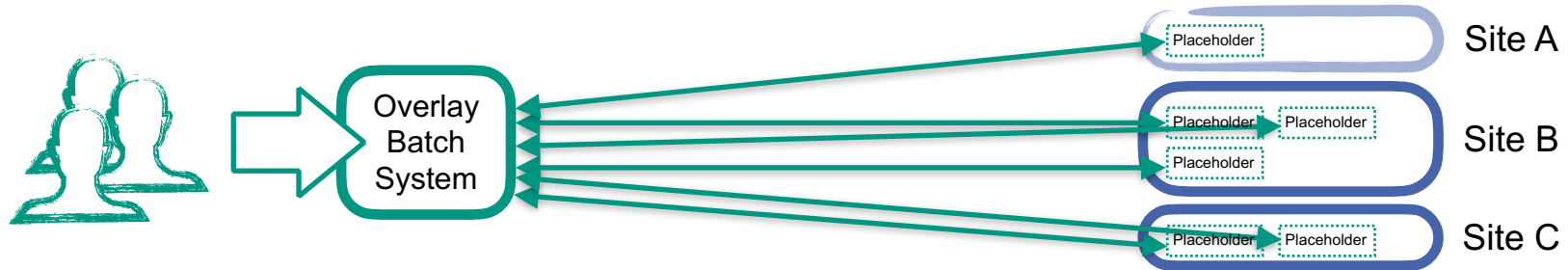


# Opportunistic Resources and their Challenges

## Opportunistic Resource

Any resources **not permanently dedicated** to but **temporarily available** for a specific task, user or group. → No reliance upon WLCG policies!

- Each opportunistic resource is different (very heterogenous system)
  - Hide complexity from users and computing operations of experiments



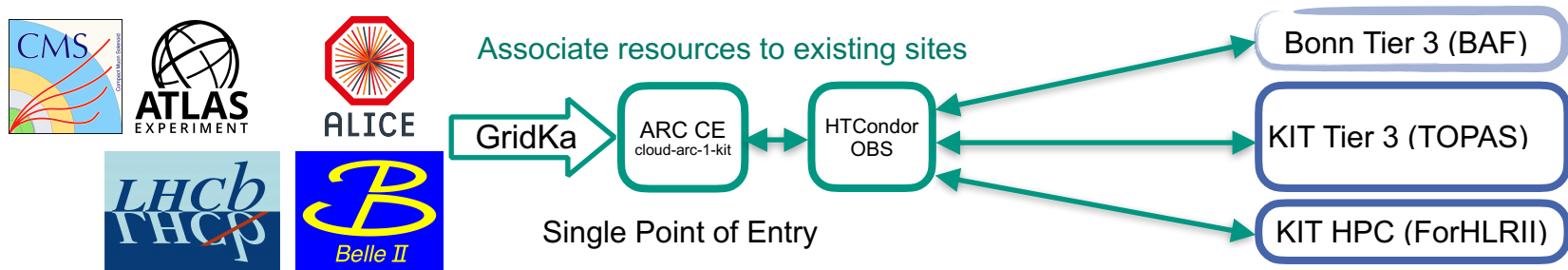
→ Dynamic and transparent integration of resources needed

# Opportunistic Resources and their Challenges

## Opportunistic Resource

Any resources **not permanently dedicated** to but **temporarily available** for a specific task, user or group. → No reliance upon WLCG policies!

- Each opportunistic resource is different (very heterogenous system)
  - Hide complexity from users and computing operations of experiments

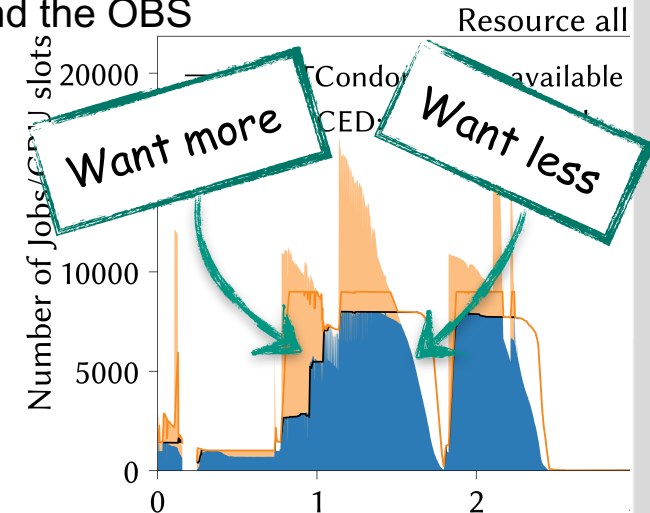
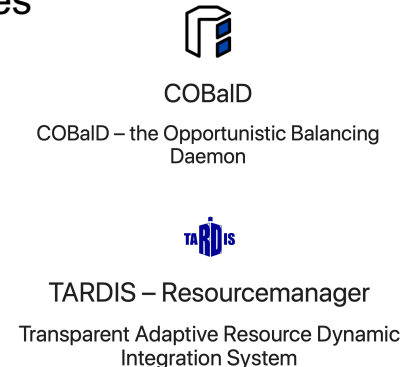


→ Dynamic and transparent integration of resources needed

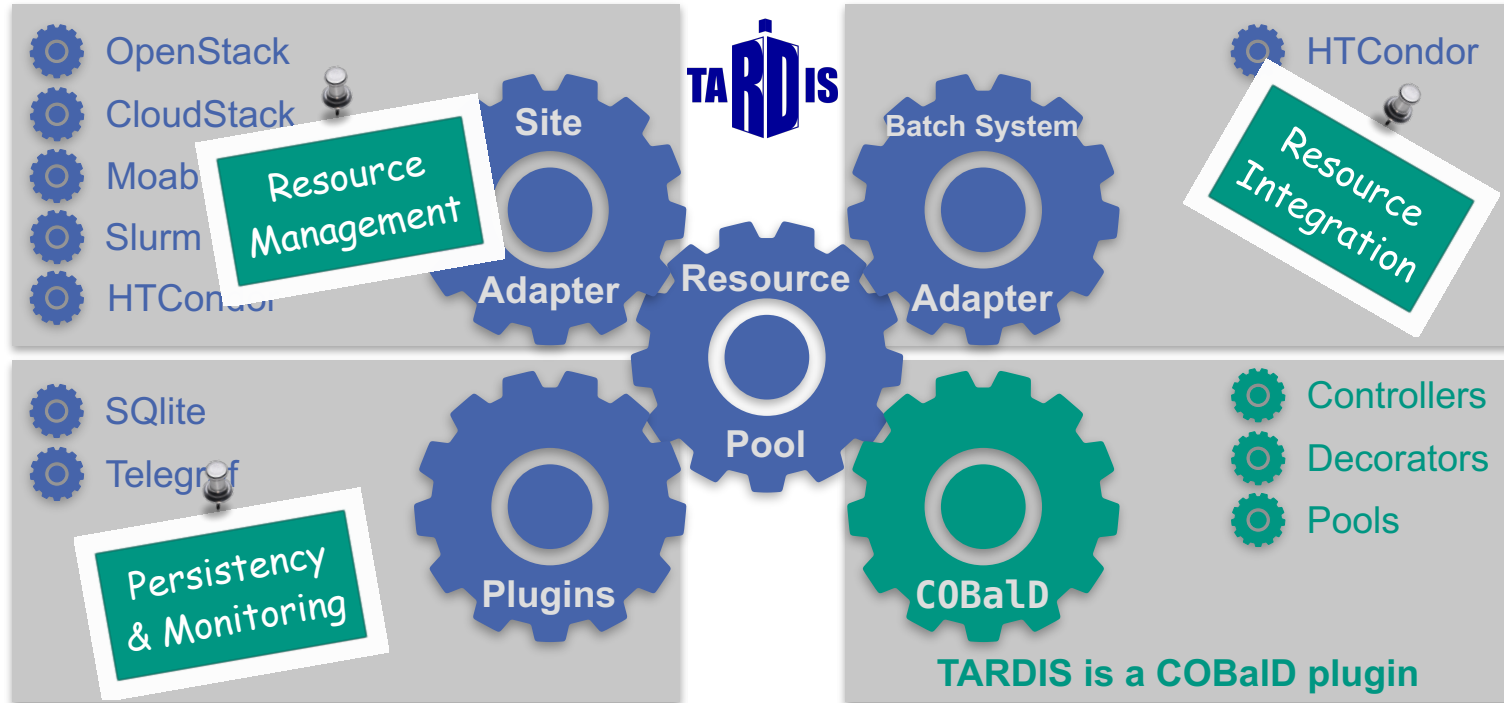
# COBaID/TARDIS Resource Manager

## Development at KIT:

- Look at what is used, not what is requested
  - Simple logic: **more used** resources, **less unused** resources
  - Regular batch system scheduler decides what to use (late binding)
  - COBaID only acquires/releases resources
    - TARDIS provides interfaces to various resource provider and the OBS
- Generic design for any resource
  - COBaID just knows (un)used resources
  - CPU, CPU+RAM, GPU, Squids, etc.
- Native composition / concurrency
  - Aggregate same resources to entire pools managed as one
  - Several COBaIDs can manage pools for same OBS (multi-agent)



# Modular Design of TARDIS



→ Easily extendable by design through its modular structure



# INITIATE AND FOSTER COLLABORATION

# ErUM Data IDT Cloud Workshops

- June 2019 (Karlsruhe)
  - SLURM overlay batch system adapter (Freiburg)
  - Light-weight integration of TIER 3 resources into the WLCG (Bonn)
- February 2020 (Aachen)
  - Light-weight operation of RWTH TIER 2 resources
  - ... and integration of close-by HPCs

Introduction into COBaldD	Max Fischer
9-1, Building 30.23	10:30 - 11:00
Dynamic Transparent Integration and Management of Opportunistic Resources with COBald/TARDIS	Manuel Giffels et al.
Lunch	
Restaurant Continent (Indian)	11:30 - 13:00
COBald & Tardis in Bonn	Oliver Freyermuth
9-1, Building 30.23	13:00 - 13:20
Tutorials: Hands-on COBald/TARDIS	
9-1, Building 30.23	14:30 - 16:00

Cloud workshops are extremely helpful to initiate fruitful collaborations!

# Collaborative Software Development

## ■ GitHub Issues & Pull Requests

- Established a review process
- Test Driven Development
- Continuous Integration ([travis-ci.org](https://travis-ci.org))
- Test Coverage ([codecov.io](https://codecov.io))

## ■ Documentation on [readthedocs.io](https://readthedocs.io)

## ■ DOI on [zenodo.org](https://zenodo.org)

## ■ Chat on Gitter ([gitter.im](https://gitter.im))

## ■ ... plus email and phone calls

Filters

is:issue is:open

Labels 10

Milestones 1

New issue

12 Open

27 Closed

Author

Label

Projects

Milestones

Assignee

Sort

1

Missing resource status in moab site adapter

#140 opened 8 days ago by rlv

1

1

[RFC] Log levels in COBald / TARDIS

#137 opened 18 days ago by olfr

1

1

Add HTCondor jdl example to documentation

#132 opened on 28 Feb by giffels

1

1

DB needs deleting when using utilization in BatchSystem

#131 opened on 28 Feb by maxfischer2781

1

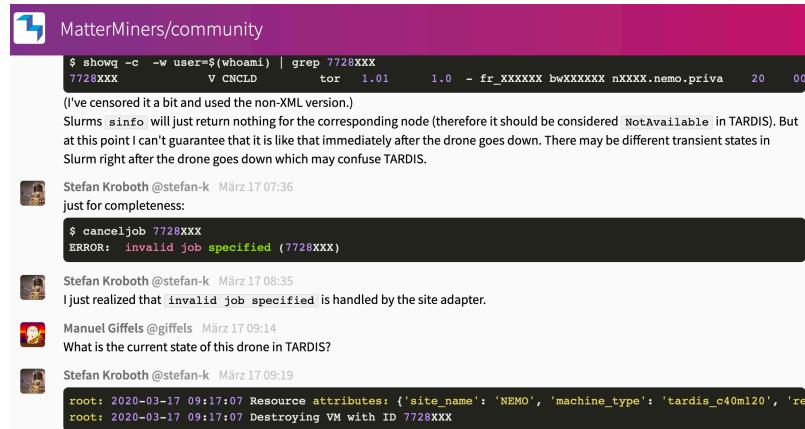
1

Job identifiers potentially non-unique in HTCondor batchsystem adapter

potential problem

#101 opened on 17 Oct 2019 by stefan-k

1



MatterMiners/community

```
$ showq -c -w user=$(whoami) | grep 7728XXX
7728XXX          V CNCLD      tor    1.01    1.0  - fr_XXXXXX bwXXXXXX nXXXX.nemo.priv  20  00
```

(I've censored it a bit and used the non-XML version.)  
Slurms `sinfo` will just return nothing for the corresponding node (therefore it should be considered `NotAvailable` in TARDIS). But at this point I can't guarantee that it is like that immediately after the drone goes down. There may be different transient states in Slurm right after the drone goes down which may confuse TARDIS.

Stefan Krobeth @stefan-k März 17 07:36  
just for completeness:  

```
$ canceljob 7728XXX
ERROR: invalid job specified (7728XXX)
```

Stefan Krobeth @stefan-k März 17 08:35  
I just realized that `invalid job specified` is handled by the site adapter.

Manuel Giffels @giffels März 17 09:14  
What is the current state of this drone in TARDIS?

Stefan Krobeth @stefan-k März 17 09:19  

```
root: 2020-03-17 09:17:07 Resource attributes: {'site_name': 'NEMO', 'machine_type': 'tardis_c40ml20', 'r
root: 2020-03-17 09:17:07 Destroying VM with ID 7728XXX
```

# Next Step: Compute Site in a Box

Unfortunately not yet funded!



## „COMPUTE SITE IN A BOX“ IN A NUTSHELL

- Develop and test new approaches to integrate tier 3 resources into a single point of entry provided by large computing centres (e. g. tier 1s)
- Fully automated deployment (provide Puppet modules for all services)
- Spread developed technology and know-how
  - Exploit results from stage 1 and roll out developed strategy on more tier 3 sites
  - foster collaboration between ErUM members and spread knowledge by means of schools and workshops
- Request 1 FTE to do the work

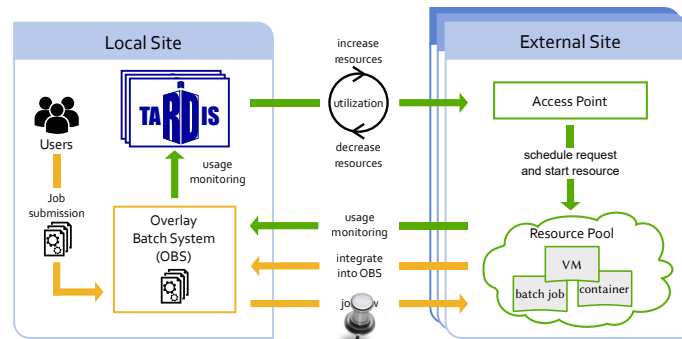
Slide by Peter Wienemann

Essential to simplify resource integration and to foster collaboration!

# ACHIEVEMENTS

# Show case: Opportunistic Compute Center for a Day

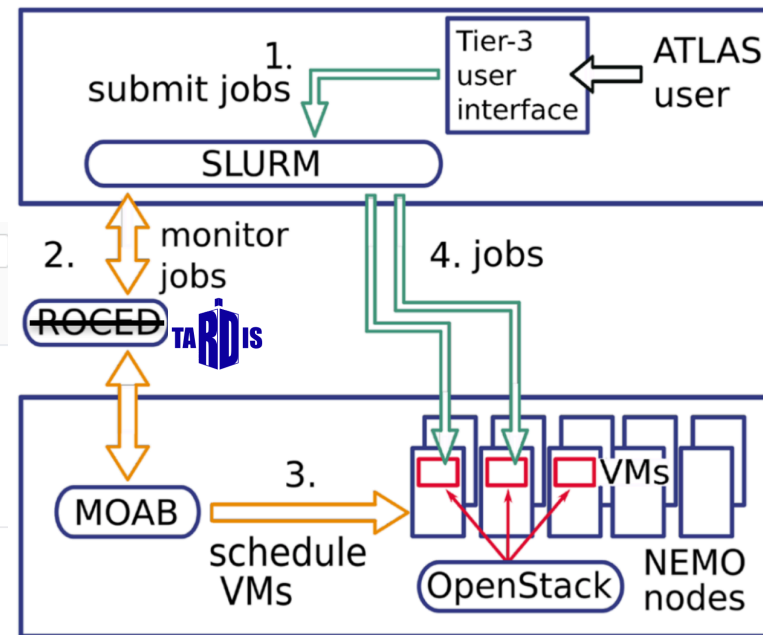
- Dynamically shared HPC centre (NEMO) at University of Freiburg (three diverse communities)
- Virtualization/containers are key components to:
  - Allow for dynamic resource integration and partitioning
  - Meet software and OS requirements
- In total around 20000 cores
  - HEP has nominal share of 33%
  - KIT has nominal share of 8%
- Transparently managed and integrated by TARDIS/C0Ba1D at ETP/KIT (TIER 3)



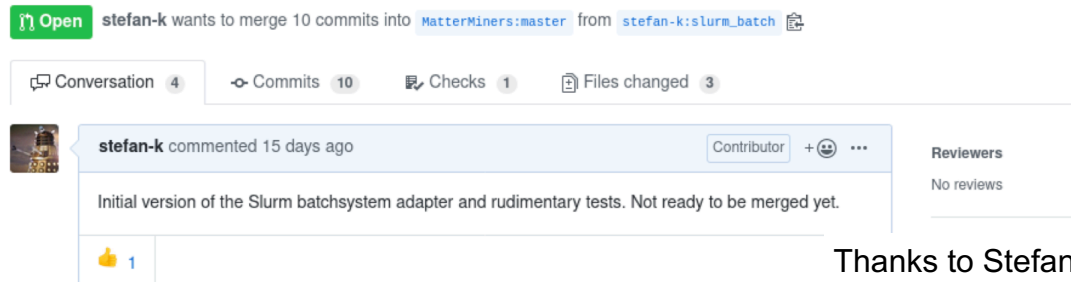
Very good experience, HEP was strongly involved in the project from the early beginning

# Dynamic Extension of the ATLAS Black Forest Grid

- Integration of HPC resources (NEMO) into ATLAS Black Forest Grid (BFG) at University of Freiburg
- Replaced ROCED with COBaLD/TARDIS at BFG
- Implemented a SLURM overlay batch system adapter for TARDIS



## WIP: Slurm batchsystem adapter #129



Thanks to Stefan Kroboth, Benoit Roland, Benjamin Rottler (U Freiburg)

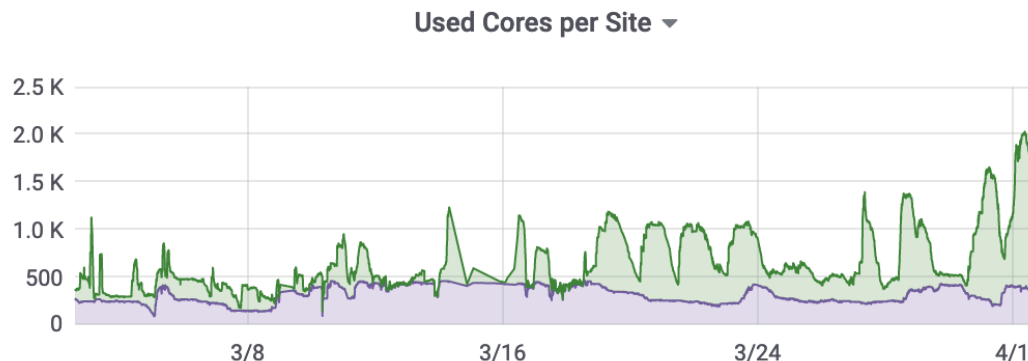
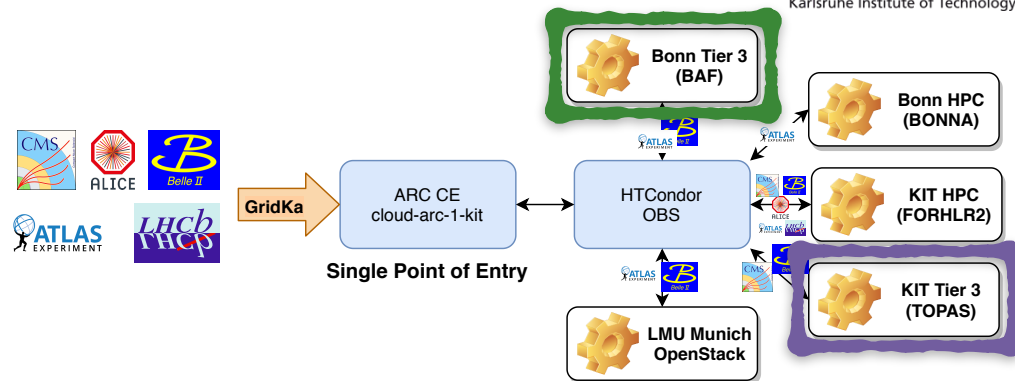
# Use Case: Backfilling of Tier 3 Resources

## In Germany:

- Local infrastructures at Universities
- Usually no Grid services deployed (lack of person power)
- Significant amount of resources
- Job requirements and fluctuations in utilization lead to unused resources

## Pilot Project @ University of Bonn:

- COBaLD/TARDIS successfully deployed @ U Bonn
- ATLAS & Belle2 production jobs are running in Bonn (Tier 3)
- Jobs are fed by an ARC-CE located at the GridKa Tier 1
- Completely transparent to the ATLAS and Belle 2 experiment



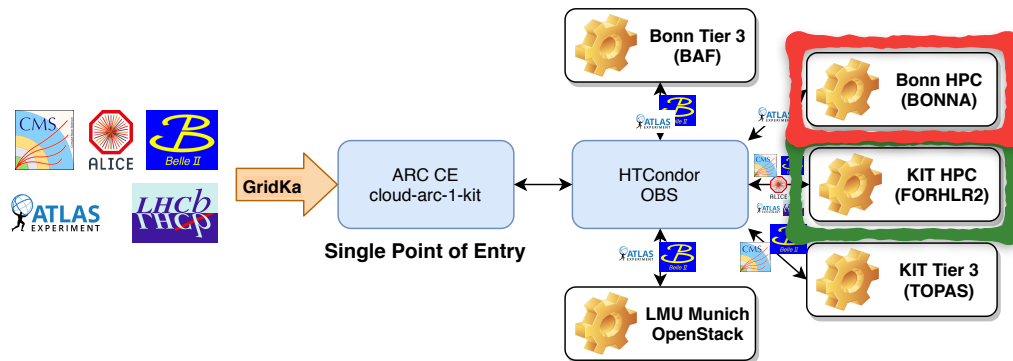
Thanks to Peter Wienemann and Oliver Freyermuth (U Bonn)



# Use Case: Integration of HPC Resources

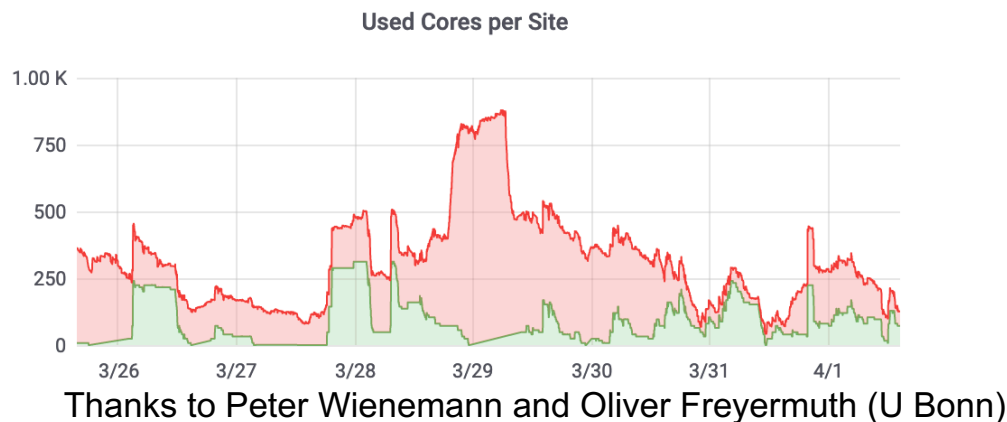
## ■ Rechencluster „Bonna“

- Fully unprivileged setup
- Charliecloud container
- Remote submission via ssh from U BONN



## ■ ForHLR2 Cluster @ KIT

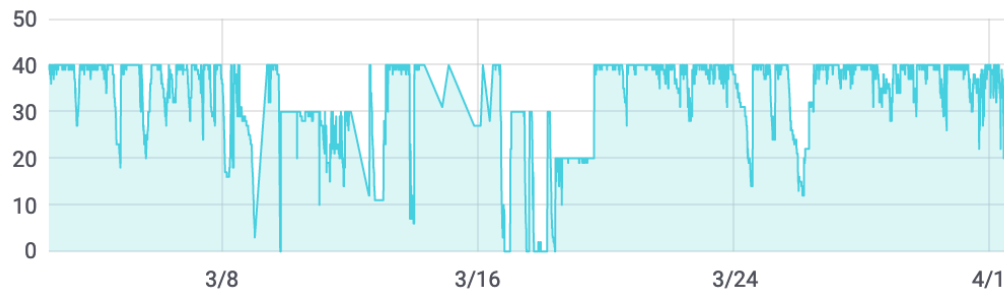
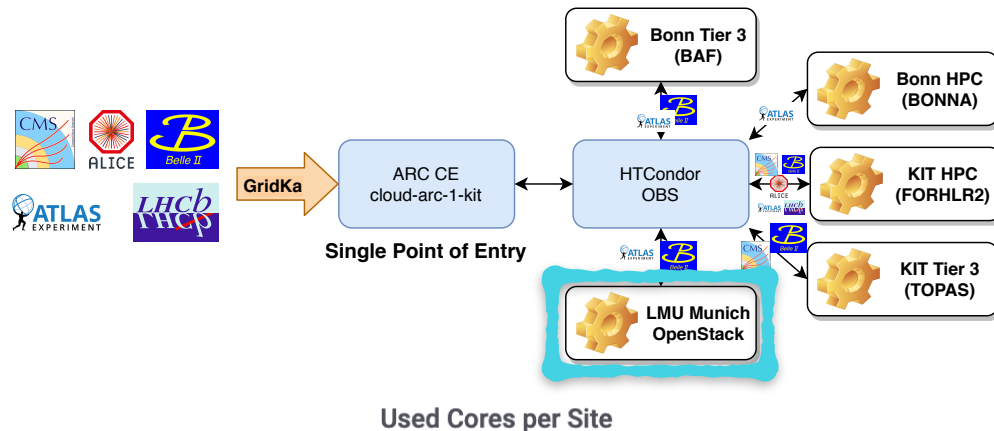
- Backfilling with HEP jobs
- Singularity container
- Remote submission via ssh from GridKa (KIT)



# Use Case: Commercial and Private Cloud Resources

## Pilot project with LMU Munich:

- C0Ba1D/TARDIS running at KIT
- Initially 40 cores assigned at the LMU OpenStack instance
- Accepts Belle2 & ATLAS WLCG jobs
- Jobs are fed by ARC-CE located at the GridKa T1
- Further projects with LMU are planned  
(C2PAP using „Site in a Box“)



Thanks to Michael Holzbock, Günter Duckeck, Rodney Walker, Thomas Kuhr!

# Use case: Light-weight Operation of Grid Resources

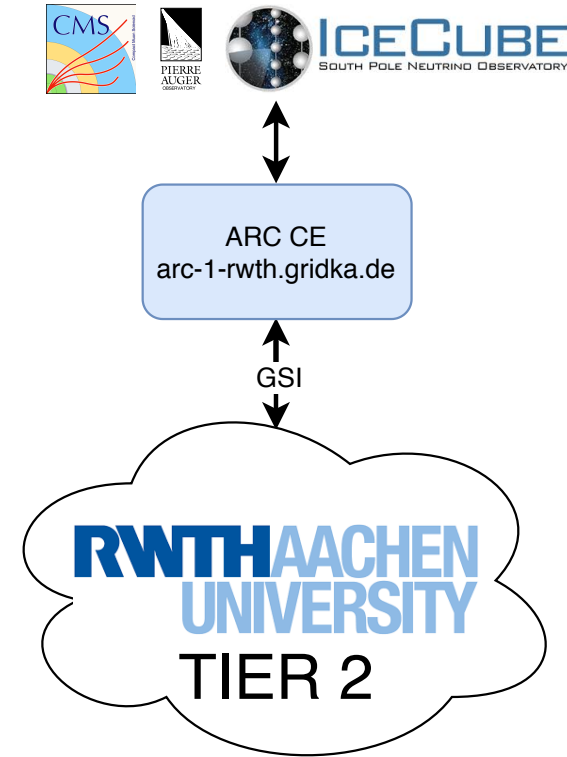
## Pilot Project with RWTH Aachen:

- Remote ARC-CE operated by GridKa managed by puppet (One of many!)
- Submit host for RWTH T2 HTCondor Test Cluster authenticated by GSI
- First test Grid jobs run successfully

## Next steps:

- Register and enable remote ARC-CE for CMS production workflows

Simplifies the operation of „smaller“ sites!



# Fight against COVID-19 with COBaLD/TARDIS

- Use COBaLD/TARDIS as job factory ensures constant job pressure
- Support of Folding@Home and Rosetta@Home
- 8 NVidia V100 + T3 backfilling at KIT-ETP
- Up to 7600k cores at GridKa T1
- ➔ Nice to see how HEP technology can help

## KIT-GridKa

Benutzer ID	2127744
Rosetta@home Mitglied seit	30 Mar 2020
Land	Germany
Gesamtguthaben	1,369,127
aktueller Punktedurchschnitt	123,597.87



## Team: KIT-ETP

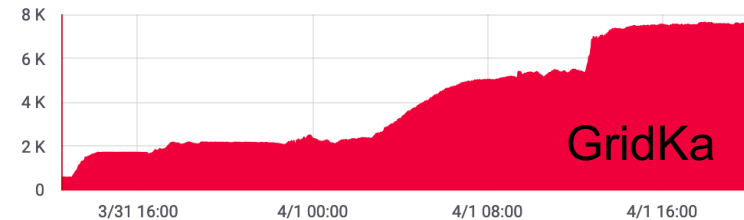
Date of last work unit 2020-04-02 10:02:40  
Active CPUs within 50 days 13,083  
Team Id 250565  
Grand Score [19,208,084](#)  
Work Unit Count [13,923](#)  
Team Ranking 3286 of 247079  
Homepage [etp.kit.edu](http://etp.kit.edu)



## Team members

Rank	Name	Credit	WUs
33,104	<a href="#">KIT-ETP</a>	11,691,266	1,492
37,716	<a href="#">GridKa</a>	7,516,818	12,431

Used Cores per VO



Give it a try! Configuration templates @ <https://github.com/MatterMiners/FoldingAtHome>

# Conclusions

- COBaID/TARDIS resource manager developed at KIT + external contributions (Bonn, Freiburg)
- Single point of entry for WLCG jobs@GridKa (dedicated ARC CE), supports multiple VOs
- Enables transparent and dynamic on-demand provisioning of opportunistic resources
- Substantial opportunistic cores provided thanks to collaborative efforts of KIT, Bonn and Munich
- Developed a concept for light-weight operation of Tier 2 compute resources (Pilot: RWTH Aachen)
- Our technology can also help to fight against COVID-19



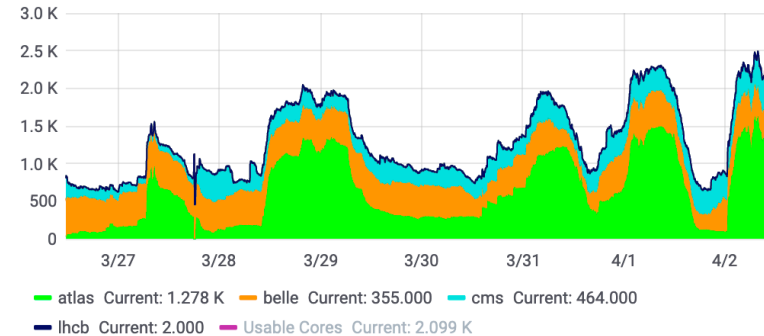
COBaID

DOI 10.5281/zenodo.3469929



DOI 10.5281/zenodo.3257718

Used Cores per VO



Used Cores per Site

