

Tools, integration, and operation of heterogeneous resources in scientific computing

Michael Böhler, Anton J. Gamel, Stefan Kroboth,
Benjamin Rottler, Markus Schumacher

IDT-ErUM collaboration meeting
February 14-15, 2022

Albert-Ludwigs-Universität Freiburg



UNI
FREIBURG

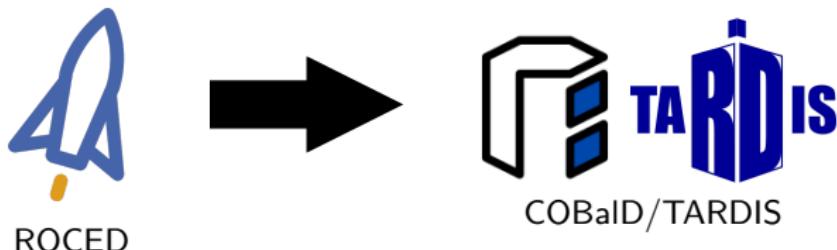
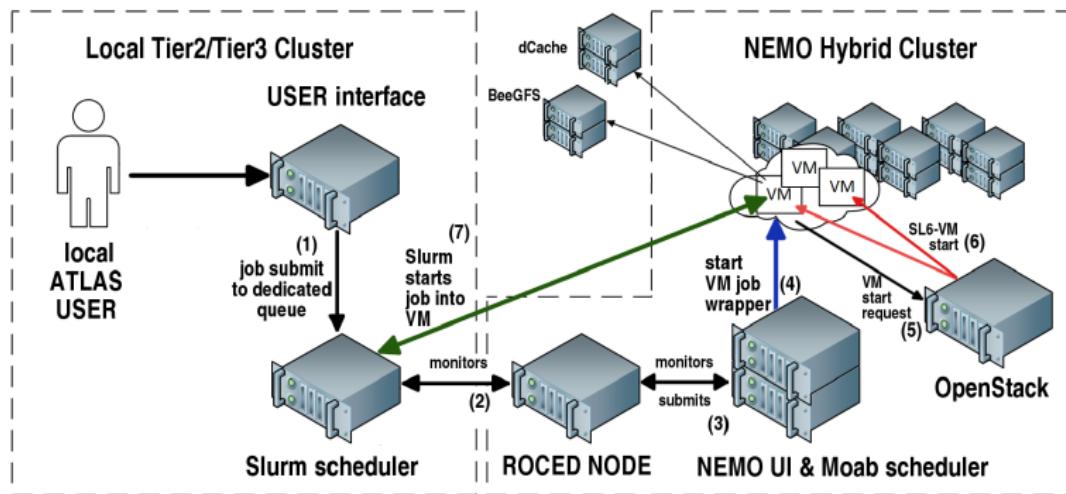
Themenbereich A: Entwicklungsarbeiten zur Bereitstellung von Technologien zur Nutzung heterogener Computing-Ressourcen

- ▶ Arbeitspaket 1: Werkzeuge zur Einbindung von heterogenen Ressourcen in das wissenschaftliche Rechnen
- ▶ Arbeitspaket 2: Effiziente Nutzung von heterogenen Ressourcen
- ▶ Arbeitspaket 3: Identifikation und Steuerung von Workflows auf heterogenen Ressourcen

Themenbereich B: Anwendung und Test von virtualisierten Softwarekomponenten im Umfeld heterogener Computing-Ressourcen

- ▶ Arbeitspaket 2: Job- und Ressourcenmanagement
- ▶ Arbeitspaket 4: Kombinierte Tests

From ROCED to COBaID/TARDIS



Contributions to COBald/TARDIS

SLURM batch system adapter

- ▶ Allows TARDIS to connect to the SLURM scheduler
- ▶ Useful for anyone running SLURM as an OBS

Monitoring plugins

- ▶ Prometheus monitoring plugin for *time series* data:
How many drones are in which state?
- ▶ Elasticsearch monitoring plugin for *event-like* data:
How and when did the drones transition between states?

Smaller contributions

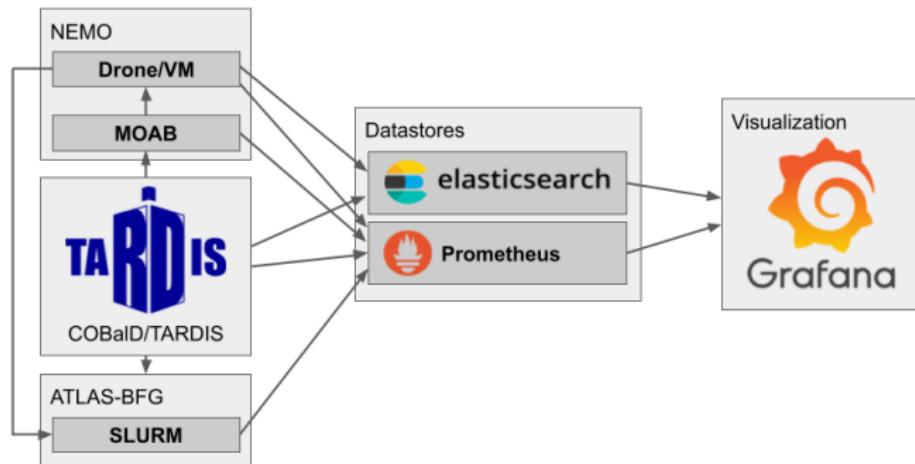
- ▶ Support executors in Slurm and HTCondor batch system adapters
- ▶ Allow users to specify additional parameters for Moabs msub command
- ▶ Bugfixes

Monitoring

- ▶ Complex infrastructure with many components and interactions
- ▶ Varying temporal dependencies
- ▶ Difficult to debug problems
- ▶ Difficult to assess the impact on performance when changes are made

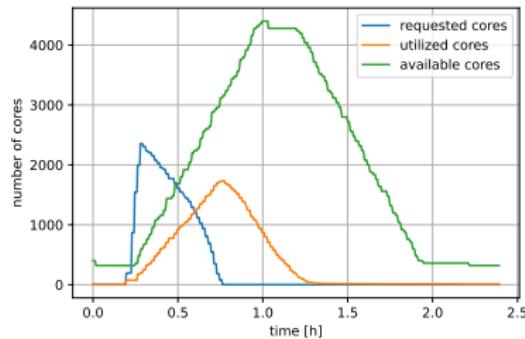
Observability

- ▶ Collect data from various sources into data storages
 - ▶ Implemented metrics collectors
- ▶ Aggregate them in a Grafana dashboard

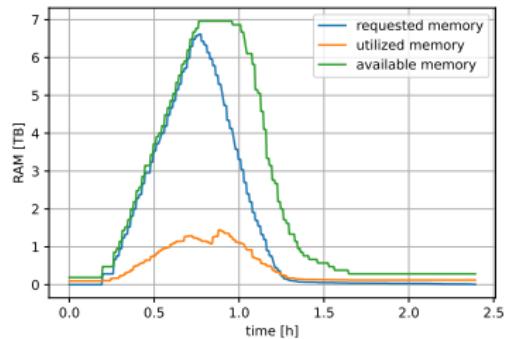


Monitoring

CPU cores



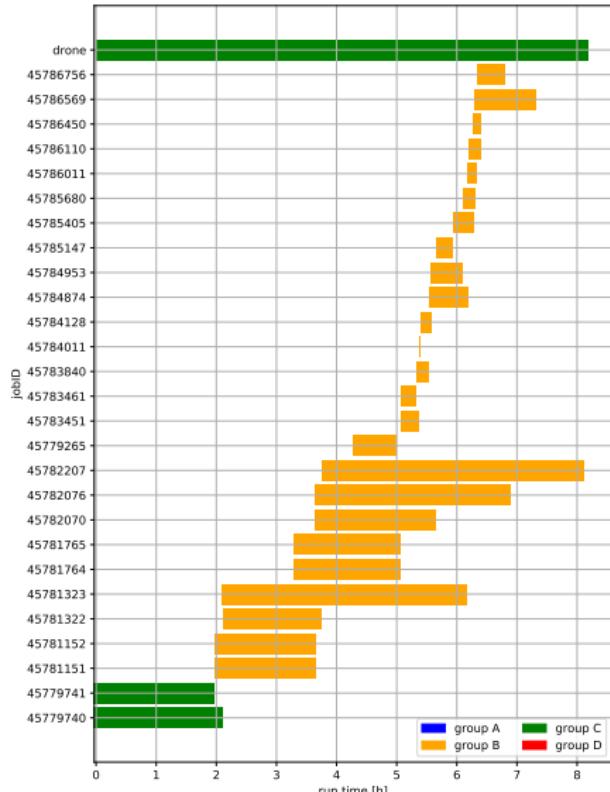
RAM



- ▶ Some CPU cores running idle because of RAM requirements of jobs
- ▶ Steepness of increase in available resources could be a measure for the responsiveness of the setup
 - ▶ Influenced by: COBald/TARDIS settings, NEMO utilization, ...

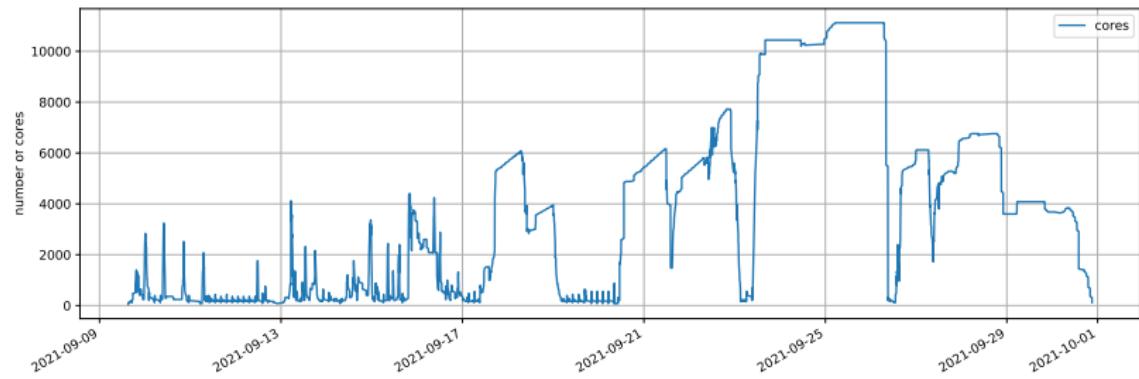
Monitoring

- ▶ Aggregating data from various sources creates additional value
- ▶ Example: Visualization of job runtime within a drone requires
 - ▶ Drone data (COBald/TARDIS)
 - ▶ Job data (SLURM)



Monitoring

Number of integrated cores



- ▶ Resources are only integrated when needed and released otherwise
- ▶ Efficient use of resources across cluster boundaries

Challenges

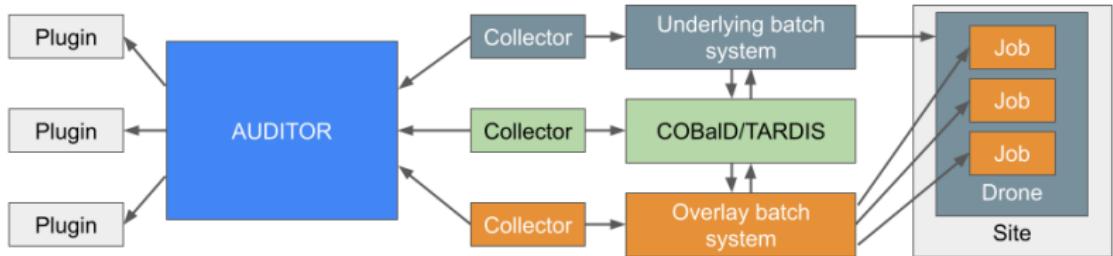
- ▶ Many potential use cases
 - ▶ Accounting with an experiment such as ATLAS/CMS/...
 - ▶ Inter-site billing between sites sharing resources
 - ▶ Matching fairshares/priority to provided resources
- ▶ Vastly different infrastructure

Flexible tracking of resource usage is vital

- ▶ Collecting data
- ▶ Storing data
- ▶ Interface to other (external) services

AUDITOR

Accounting Data handling Toolbox for Opportunistic Resources



- ▶ AUDITOR: Accepts information about consumed resources and stores it in a database
- ▶ Collector: Obtains information from a source (Batch system schedulers, COBald/TARDIS, ...) and forwards it to AUDITOR
- ▶ Plugin: Fetches data from AUDITOR and takes an action

- ▶ Main functionality implemented (accepting, storing and providing data)
- ▶ AUDITOR prototype
 - ▶ Implemented in Rust
 - ▶ Actix-web actor framework
 - ▶ REST interface
 - ▶ PostgreSQL database (other data backends possible)
 - ▶ Completely stateless
- ▶ Plugins/Collectors
 - ▶ Framework for interaction with AUDITOR (in Python)
 - ▶ TARDIS accounting plugin (based on monitoring interface)
- ▶ Open points
 - ▶ Example usecase
 - ▶ Authentication



Summary

- ▶ Opportunistic resources
 - ▶ Switch to COBald/TARDIS
 - ▶ Supports different kinds of resources
 - ▶ Automatically decides which to integrate based on utilization
 - ▶ Necessary works:
 - ▶ SLURM batch system adapter
 - ▶ Adjust infrastructure
- ▶ Monitoring
 - ▶ Added new features
 - ▶ Implemented Prometheus and Elasticsearch monitoring plugins for COBald/TARDIS
 - ▶ Infrastructure based on open-source tools with common interfaces
 - ▶ Finetuning of job and resource management
 - ▶ Adjusted COBald/TARDIS parameters based on data from monitoring system
- ▶ Accounting
 - ▶ Prototype implemented
 - ▶ Prototype collector and plugin framework implemented
 - ▶ Implementation of fairshare

Related publications and talks

Opportunistic extension of a local compute cluster with NEMO resources for HEP workflows

Benjamin Rottler, et.al., 14th Annual Meeting of the Helmholtz Alliance "Physics at the Terascale", November 2021

Opportunistic extension of a local compute cluster with NEMO resources for HEP workflows

Stefan Kroboth, et.al., 7th bwHPC Symposium, November 2021

Transparent Integration of Opportunistic Resources into the WLCG Compute Infrastructure

M. Böhler et al., 25th Int. Conference on Computing in High-Energy and Nuclear Physics, 2021

Accounting of opportunistic resources

Stefan Kroboth, et.al., IDT-ErUM Collaboration Meeting, May 2021

Performance monitoring of the opportunistic resource NEMO at ATLAS-BFG

Stefan Kroboth, et.al., DPG-Frühjahrstagung 2021, Dortmund, March 2021

Performance Monitoring of Opportunistic Resources at ATLAS-BFG

Stefan Kroboth, et.al., Kollaborationstreffen IDT-ErUM, September 2020