Karlsruher Institut für Technologie

# On-demand provisioning of HEP compute resources on cloud sites and shared HPC centers

**Günther Erli, Frank Fischer, Georg Fleig, Manuel Giffels, Thomas Hauth, Günter Quast, Matthias Schnepf (IEKP), Andreas Petzold (SCC)**

STEINBUCH CENTRE FOR COMPUTING - SCC

# Who’s done all the work

- Institute for Experimental Nuclear Physics CMS Group

  Günther Erli
  Frank Fischer
  Georg Fleig
  Manuel Giffels
  Thomas Hauth
  Günter Quast
  Matthias Schnepf

# Motivation

- Extend resources for local users beyond the basement cluster
- GridKa Tier-1 12km away
  - Dedicated CPU and storage resources for German CMS users, available via Grid
  - OpenStack test bed
- HPC cluster with pledged resources for HEP in Freiburg 150km away
  - Local batch system + OpenStack
- Commercial cloud resources
  - 1&1
  - AWS

Tie all resources together

- Transparently
- Dynamically

# Tools

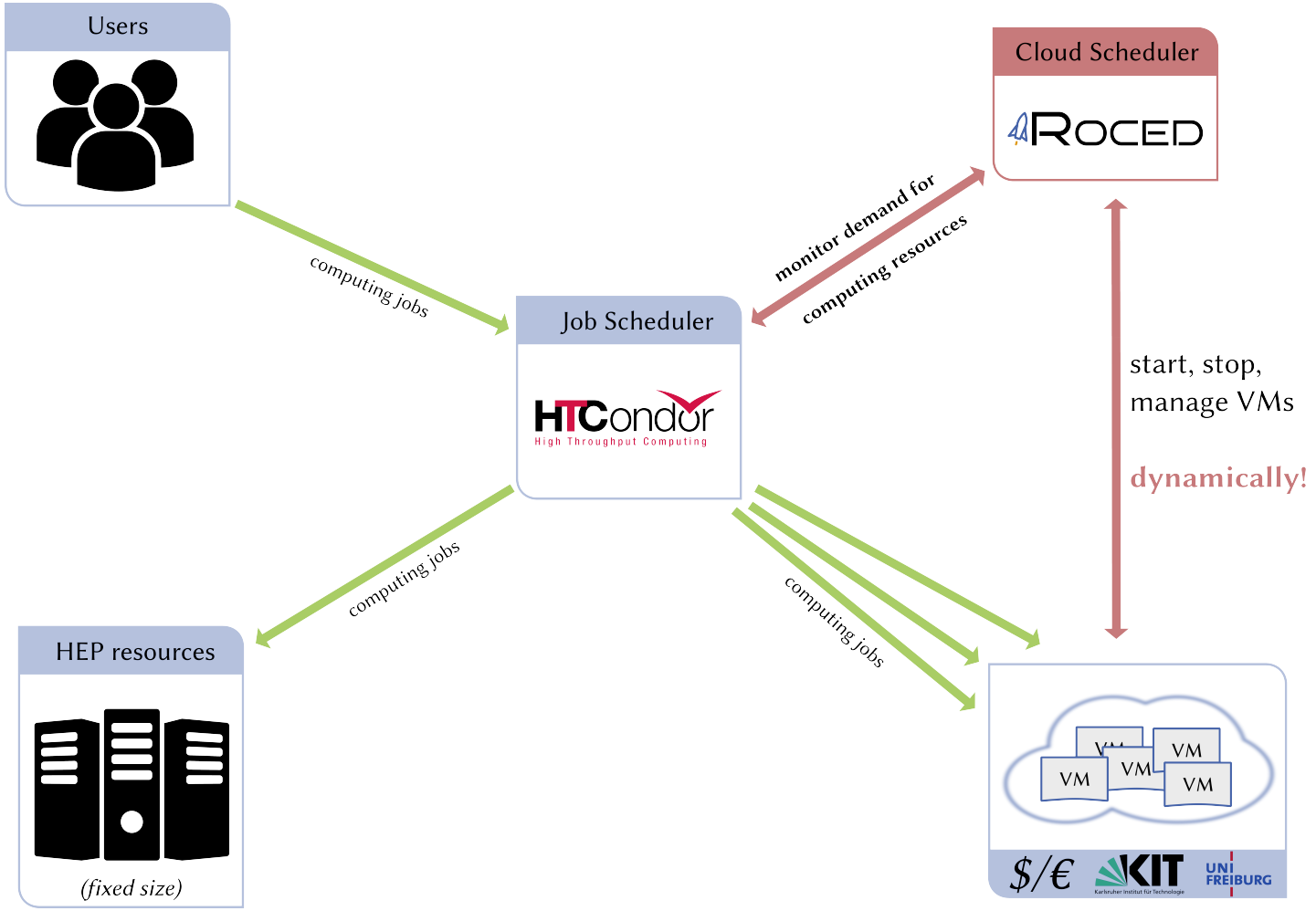
- HTCondor

  ...

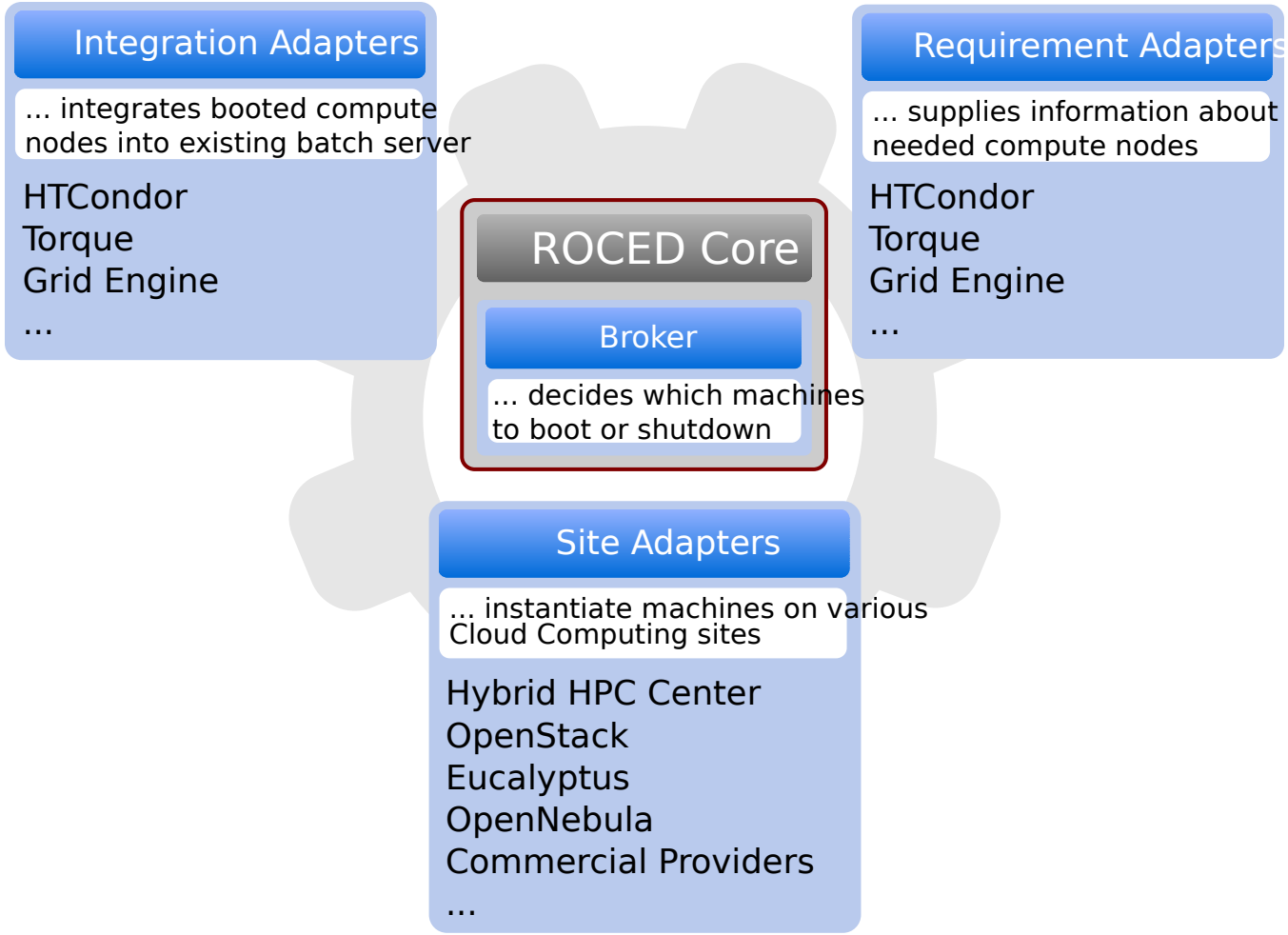
- ROCED
  - Cloud scheduler
  - Support for multiple cloud APIs and batch systems
  - Modular, easily extendable, Python
  - Parses HTCondor's ClassAds and boots VMs on cloud sites

https://github.com/roced-scheduler/ROCED

# Implementation



# *R*esponsive *O*n-demand *C*loud *E*nabled *D*eployment (ROCED)

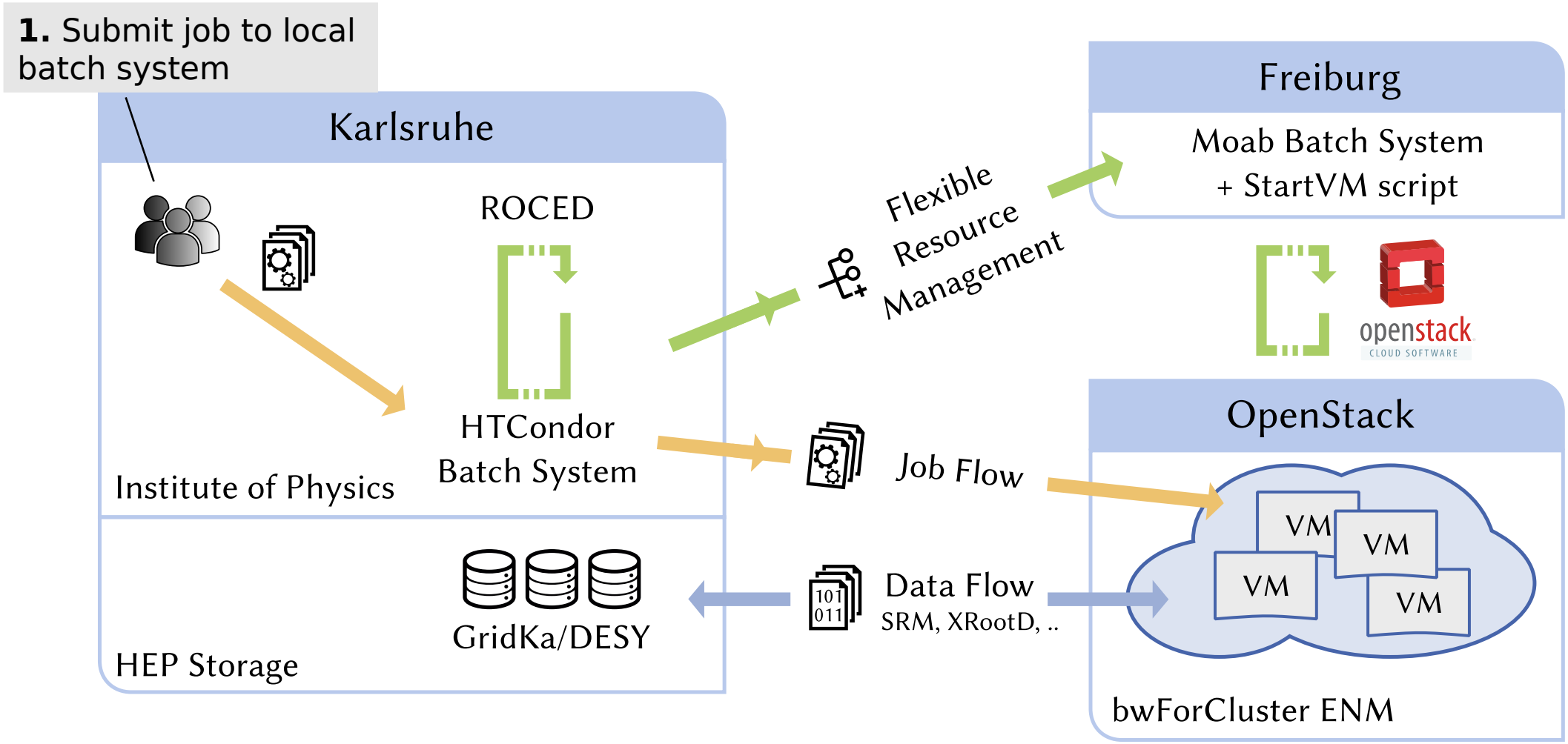


# bwForCluster NEMO @ Uni Freiburg

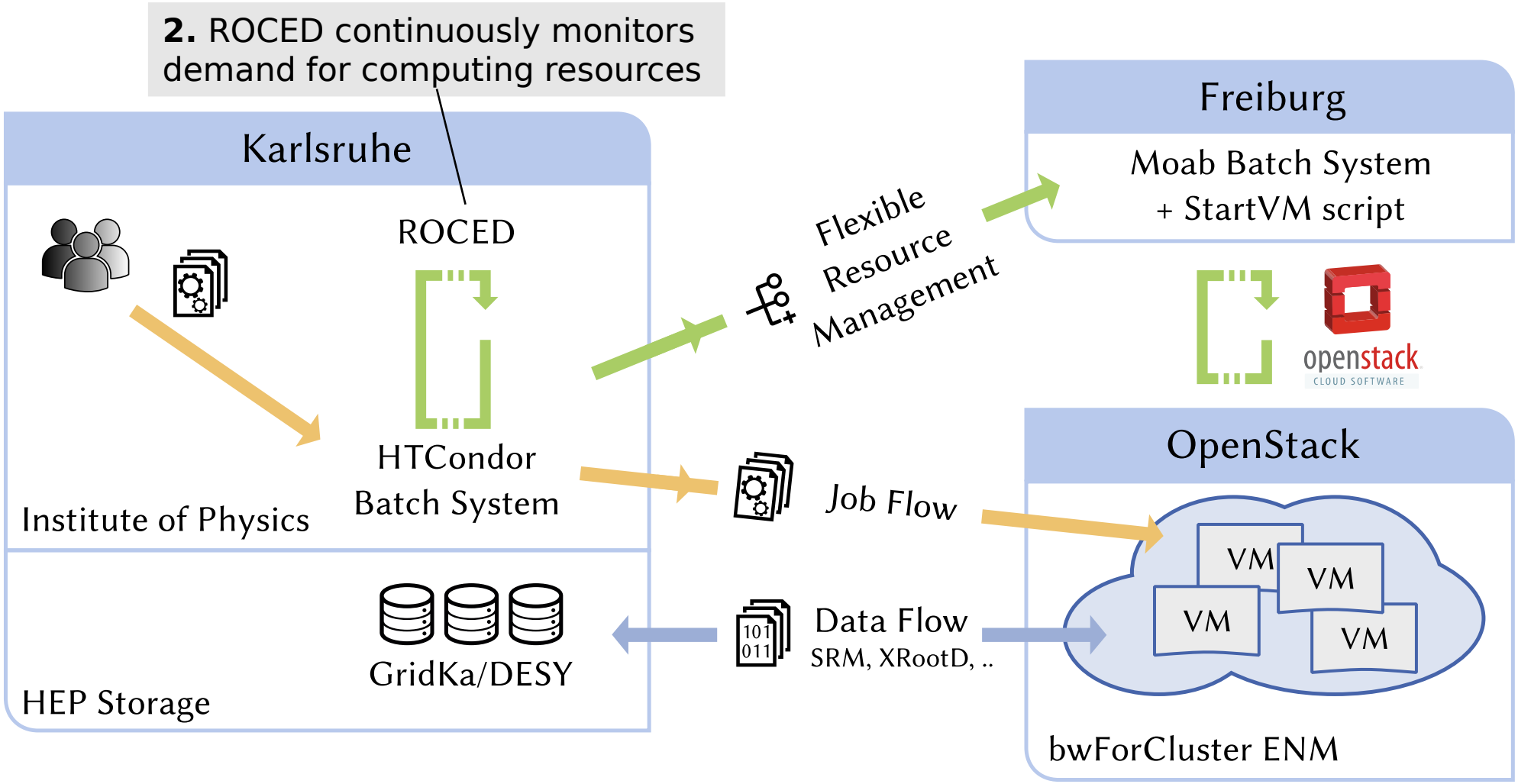
- HPC Cluster
  - available since August 2016 with 15120 cores (Broadwell)
  - 100Gb/s Omnipath, BeeGFS
  - 10Gb/s WAN connection to GridKa
- Shared by 3 diverse scientific user groups:
  - **N**euroscience, **E**lementary Particle Physics, **M**icrosystem Engineering
- Virtualization concept included from the beginning!
  - VMs run as regular batch jobs (transparent to the HPC scheduler)
    - ROCED as “translator” between HTCondor and MOAB
    - Integrate seamlessly with other users' bare metal jobs
    - Fairshare between VMs and regular jobs

https://www.hpc.uni-freiburg.de/nemo

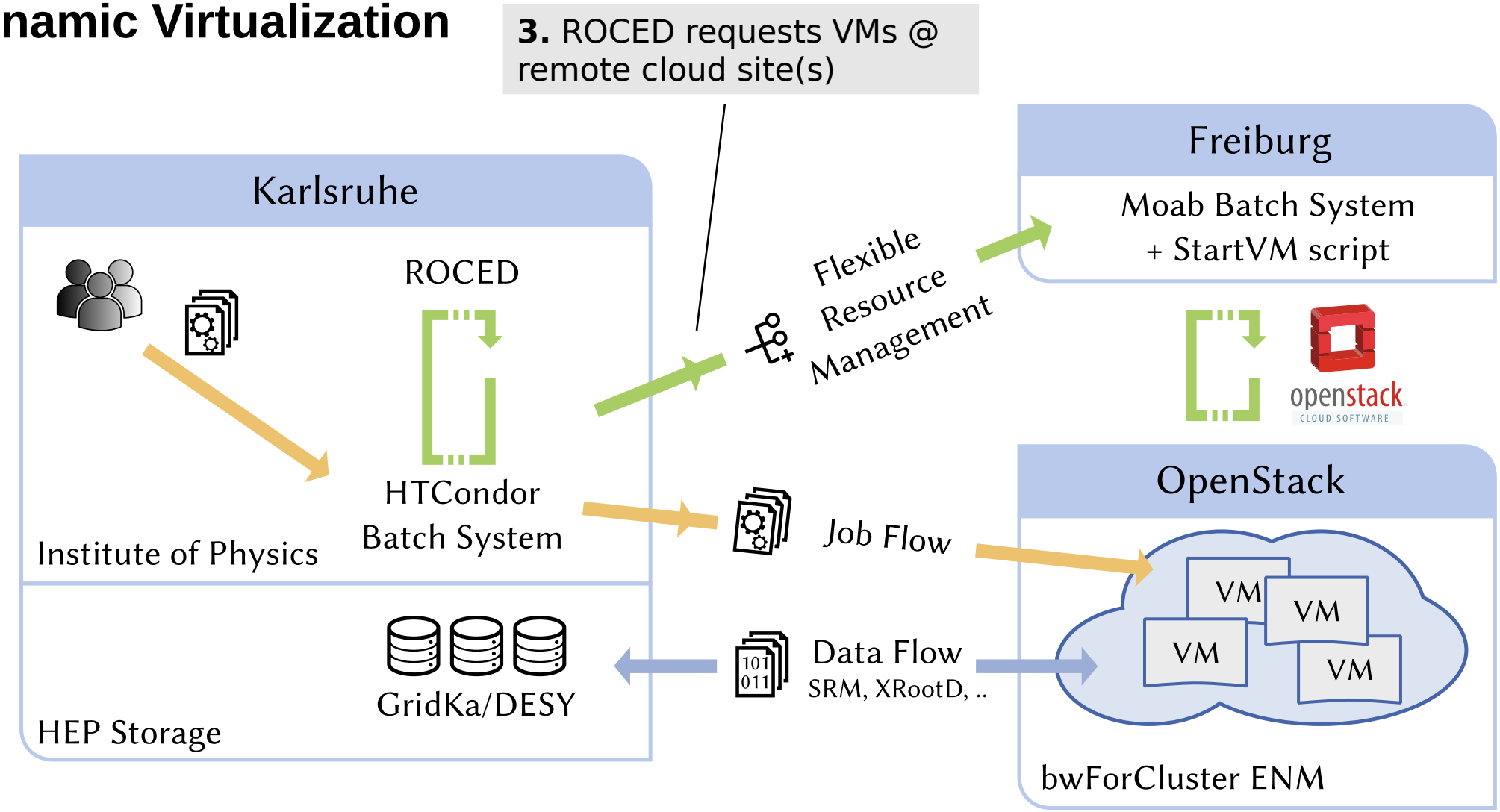
# Dynamic Virtualization



# Dynamic Virtualization

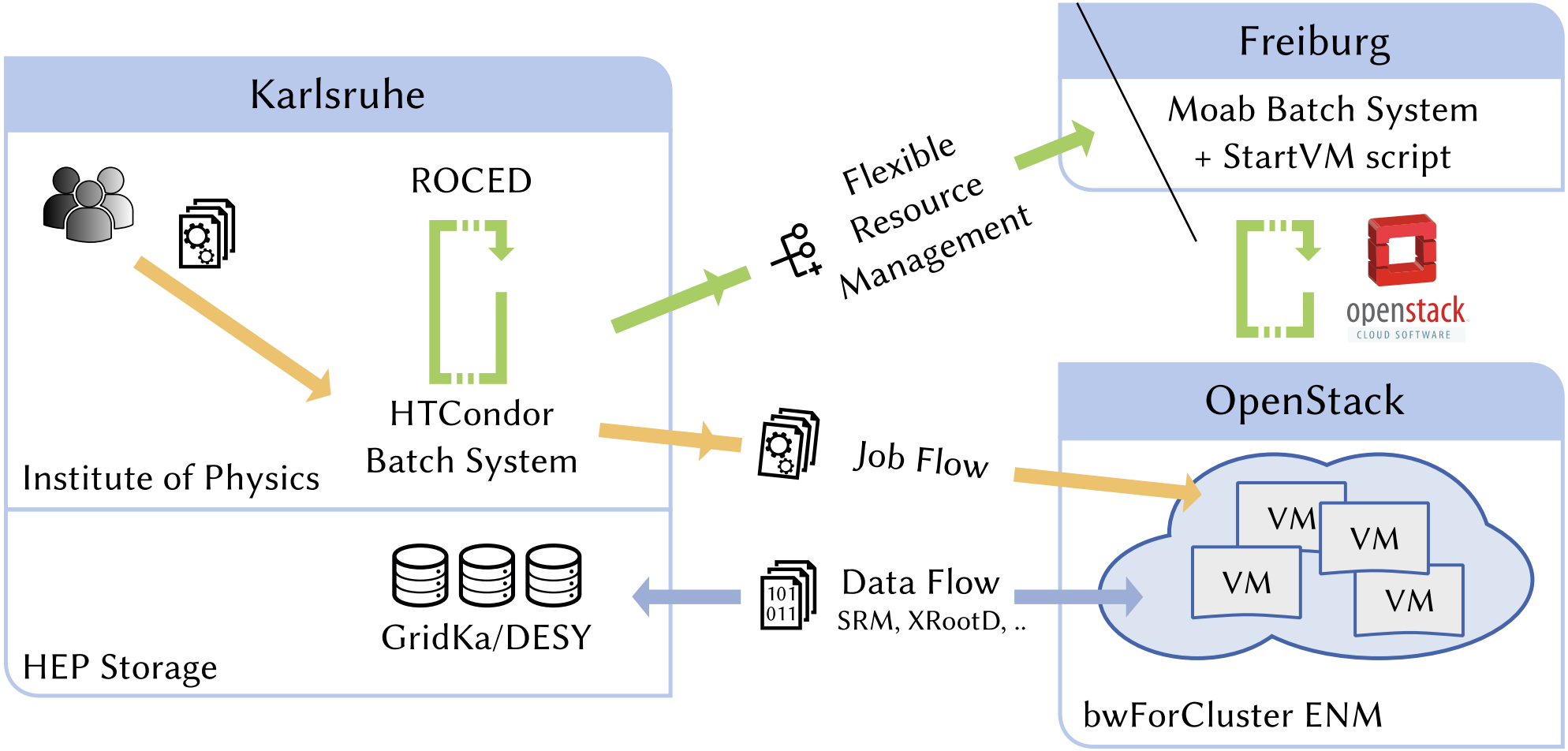


# Dynamic Virtualization

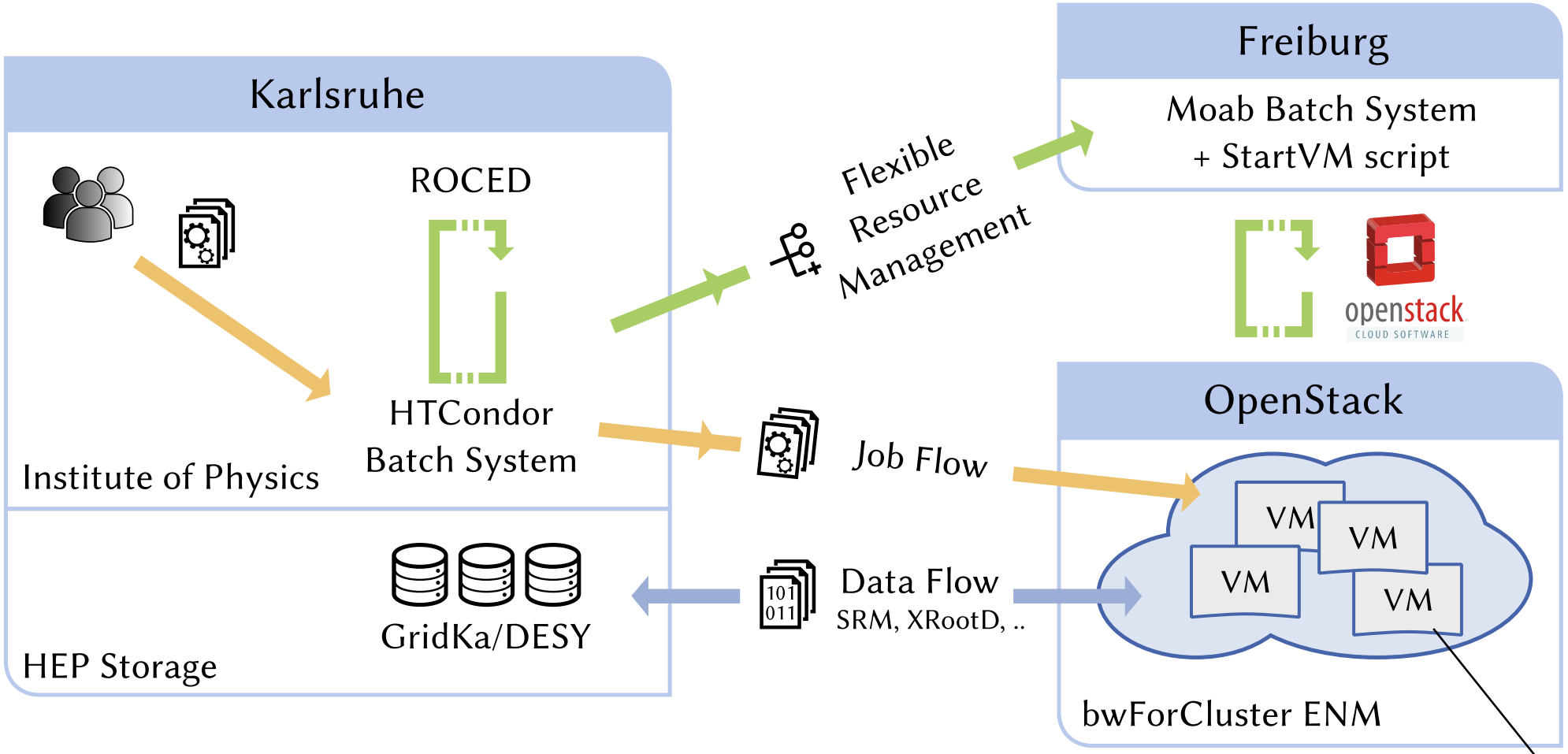


# Dynamic Virtualization

**4.** VMs start at remote site. Communication to ROCED via different APIs

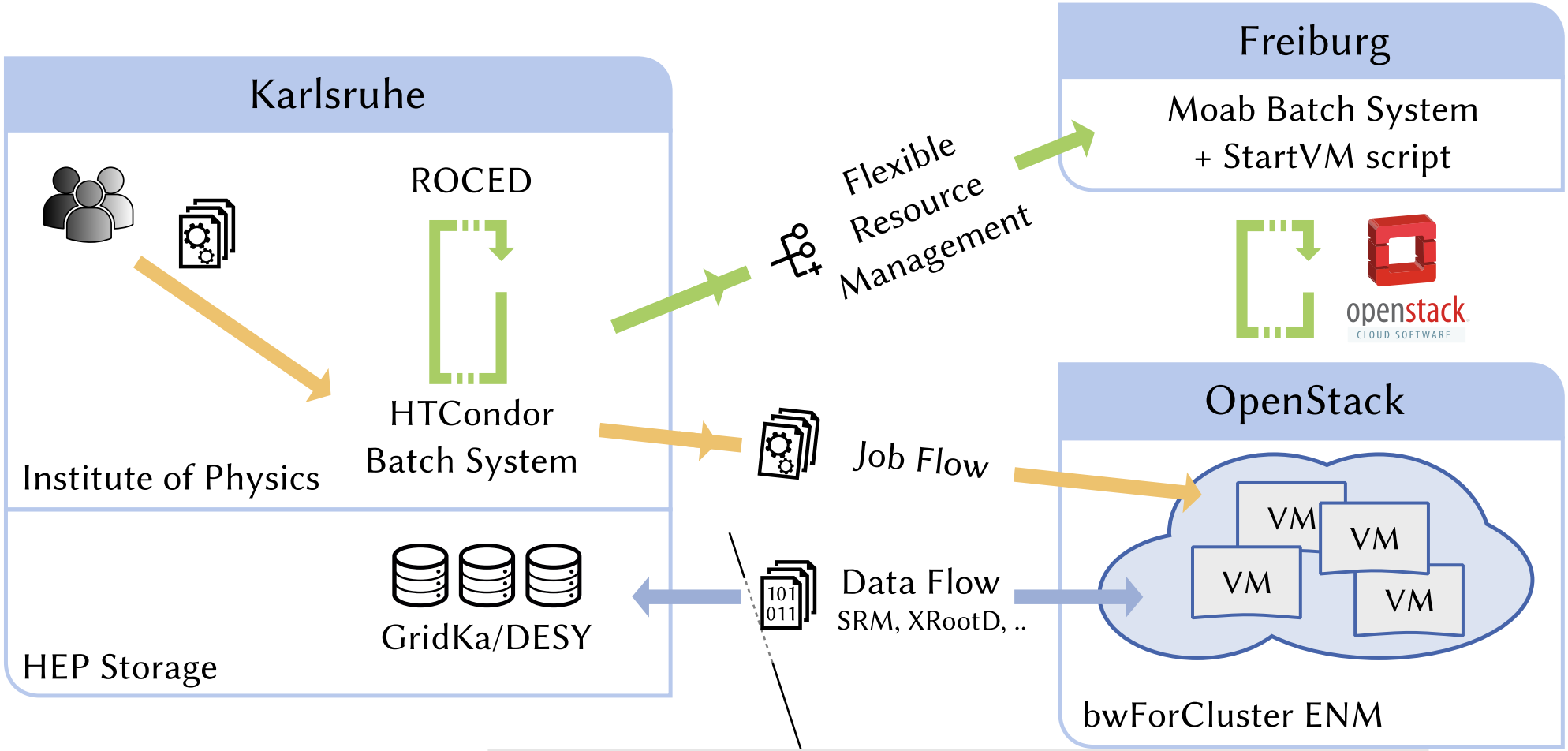


# Dynamic Virtualization



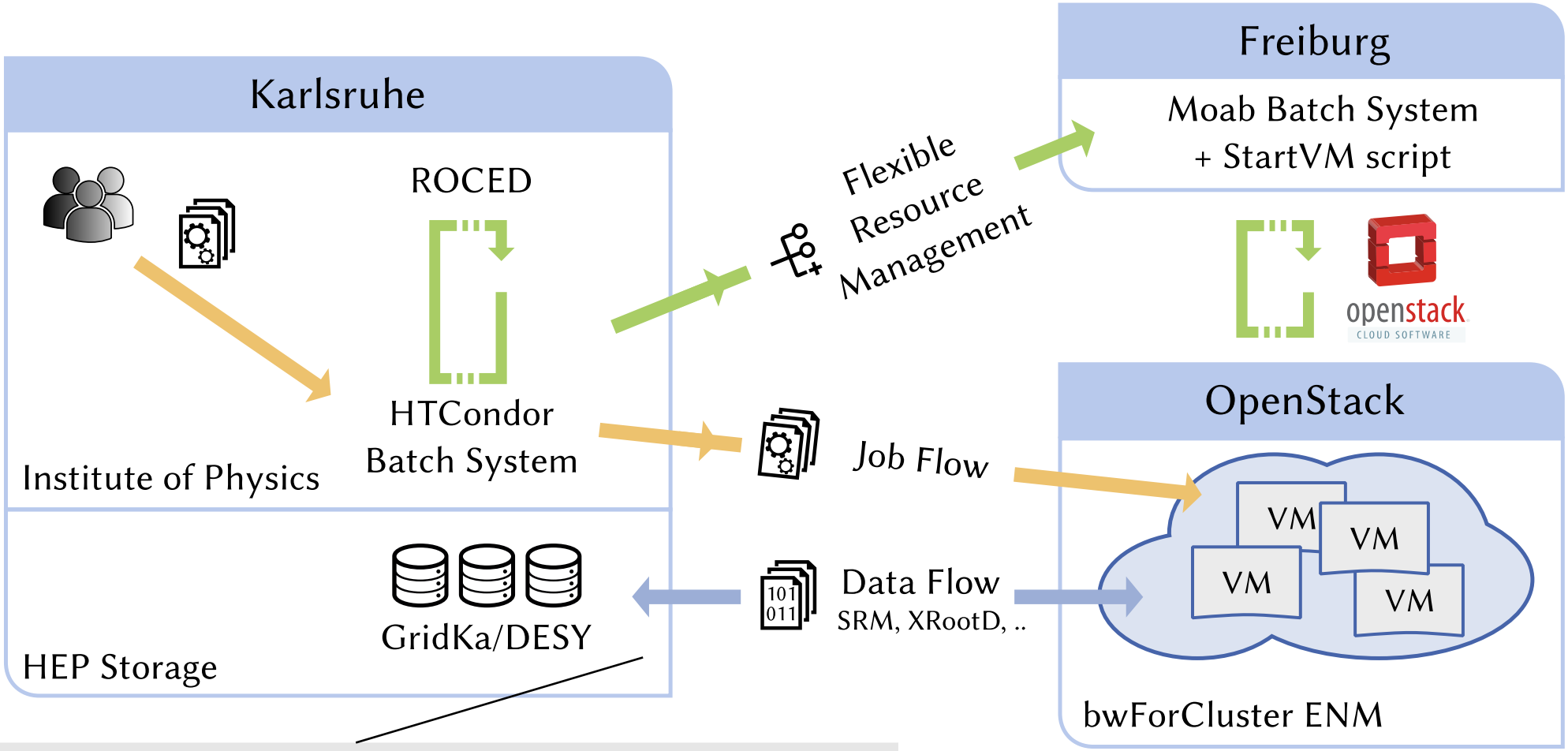

**5.** VM integrates into HTCondor

# Dynamic Virtualization



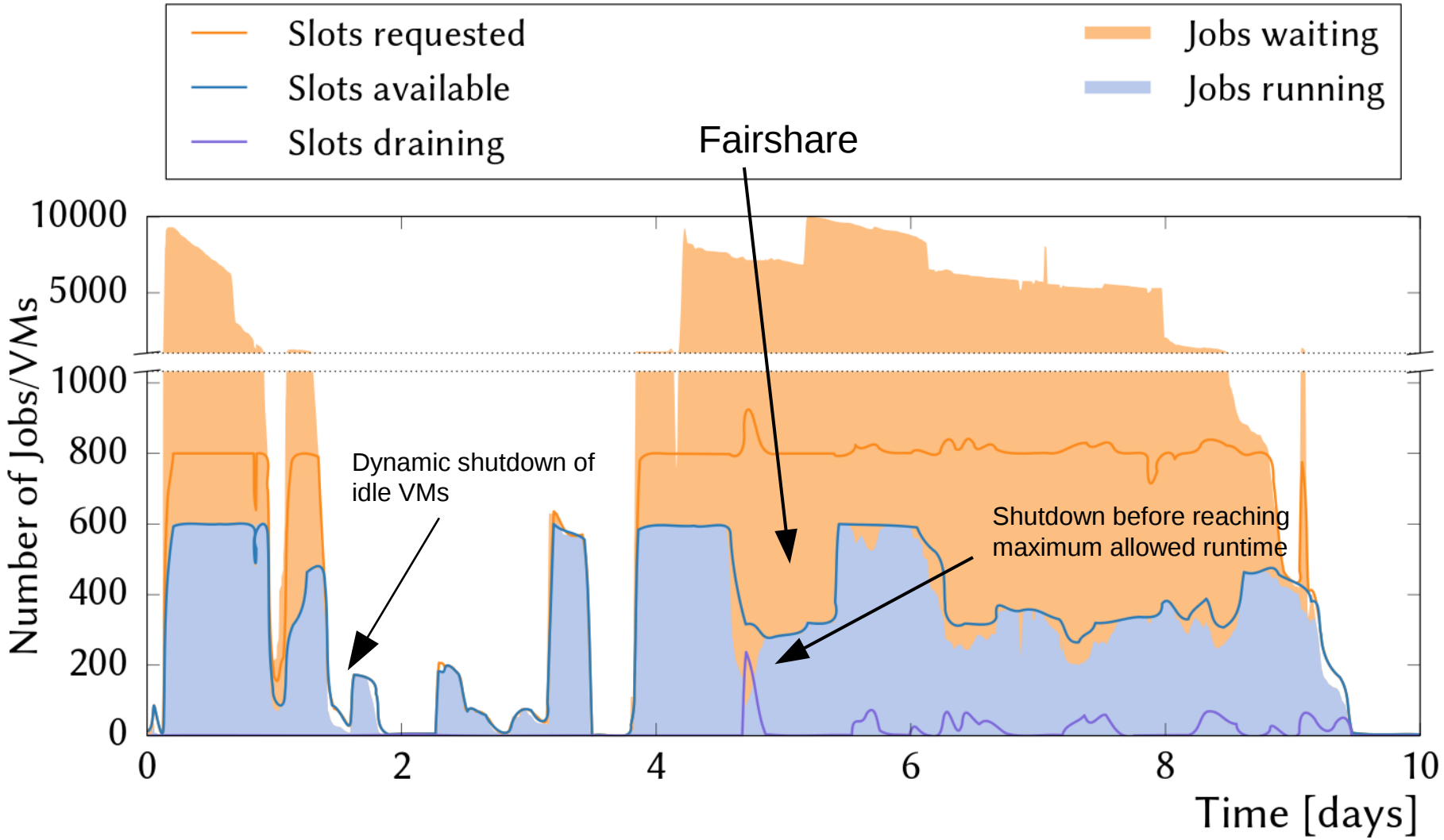

**6.** Jobs get scheduled on virtualized HEP worker node

# Dynamic Virtualization

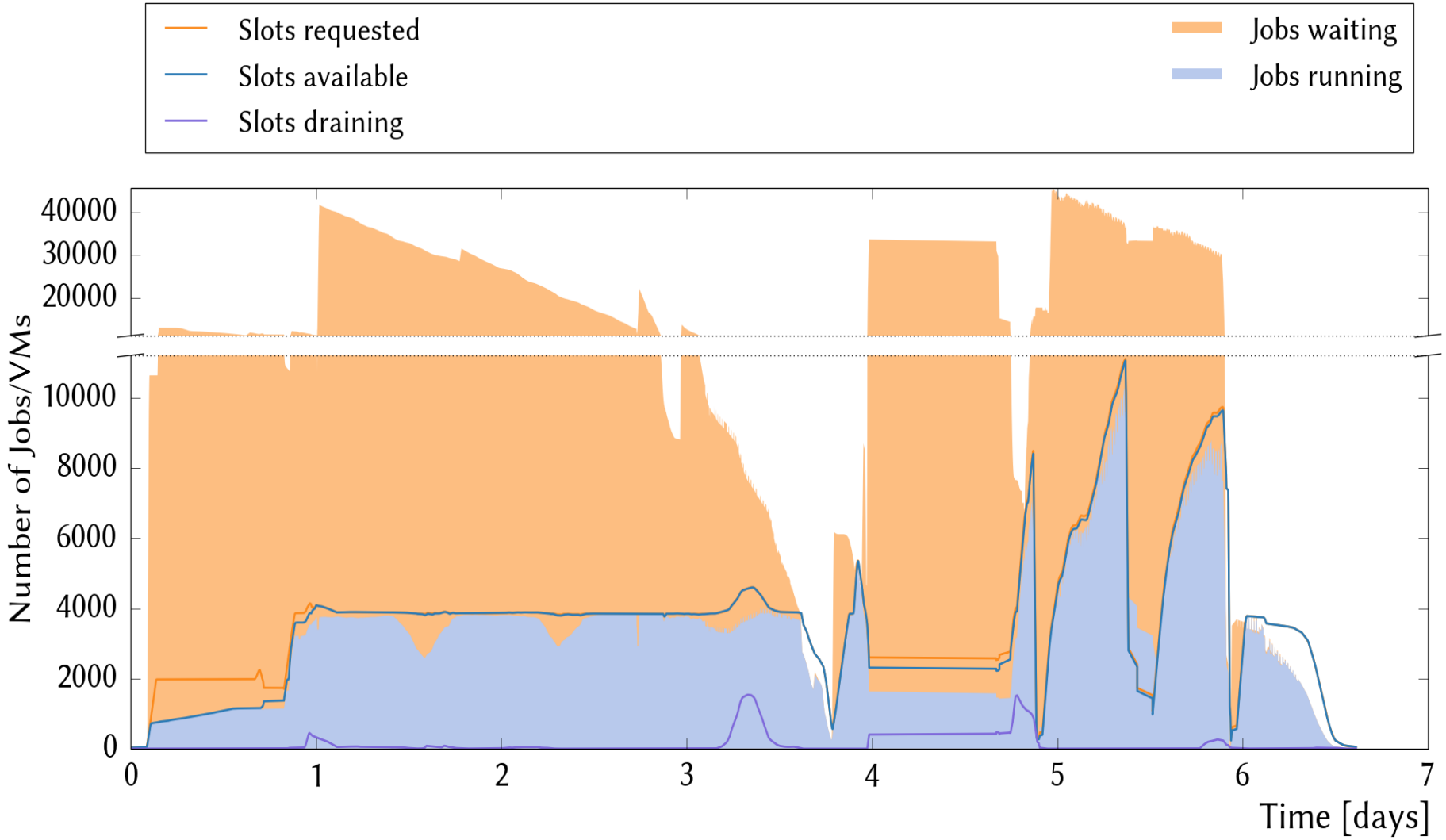



**7.** Input and output from/to HEP storage (e.g. GridKa)

# Long Term Usage (pre-production cluster)



# Dynamic “1-day Tier-1”

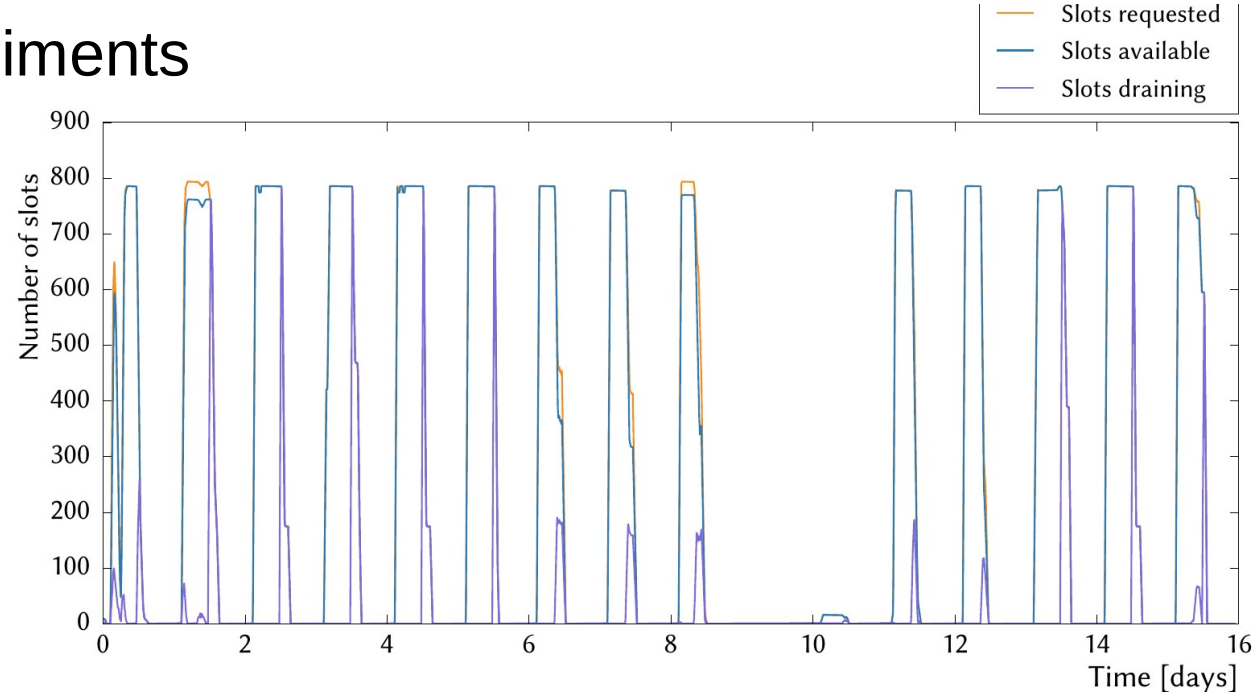


# Commercial Cloud: 1&1

- Evaluation of 1&1 Cloud Server product for HEP jobs
  - Joint 1&1 and KIT team
- ROCED adapter for Cloud Server API
- Possibility to upload and deploy a custom VM-Image with Scientic Linux 6.7 (or other, if needed) and CVMFS and HTCondor support
- Configured dedicated CVMFS-Squid VM in 1&1 data center which gets automatically started by ROCED as soon as one worker VM is booted

# Tests in 1&1 Cloud

- Take advantage of free CPU cycles during night
  - Only jobs with runtime <12h scheduled
- Up to 800 job slots provisions per night
- API based scheduling w/o manual intervention
- Jobs from multiple different experiments ran reliably in 1&1 cloud VMs



# Summary & Outlook

- Dynamic resource integration with ROCED and HTCondor very successful
- HPC & Virtualization work well together
- Docker integration also working (desktop cloud)
- Network bandwidth to remote storage crucial
  - Remote storage needs to be able to take the extra load!

- ROCED development continues
- GridKa Tier-1 is switching to HTCondor
- Plan to use ROCED for dynamic Tier-1 expansion
  - HNSciCloud