



Leibniz-Institut für
Astrophysik Potsdam

Compute Cloud

Constructing a high available container orchestration infrastructure

In connection with different AI applications

E. Tom L. Strauß, Dr. Arman Khalatyan / AG2025 / 16. September 2025

Table of Contents

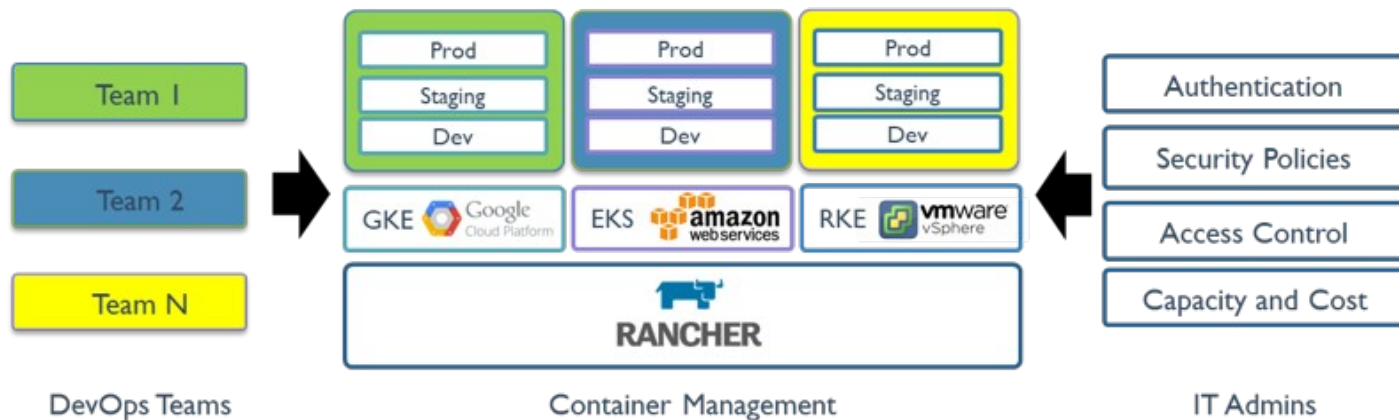
Container Orchestration Infrastructure

1. Rancher and RKE2
2. GitLab CI/CD
3. REANA (example service)

AI – Tech-Stack

4. Goals / Purposes
5. Users View
6. Integrations / Developments
7. Outlook
8. Summary

1. Rancher / RKE2



K3S



RKE



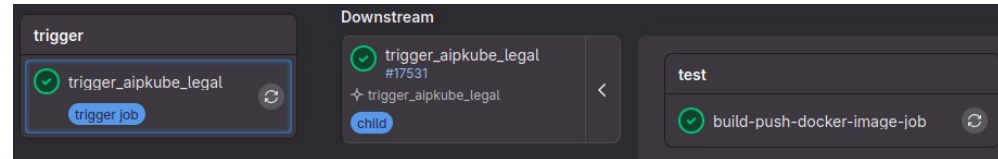
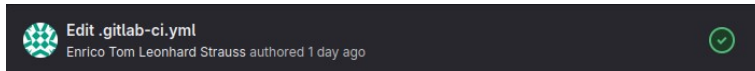
FLEET

2. GitLab CI/CD

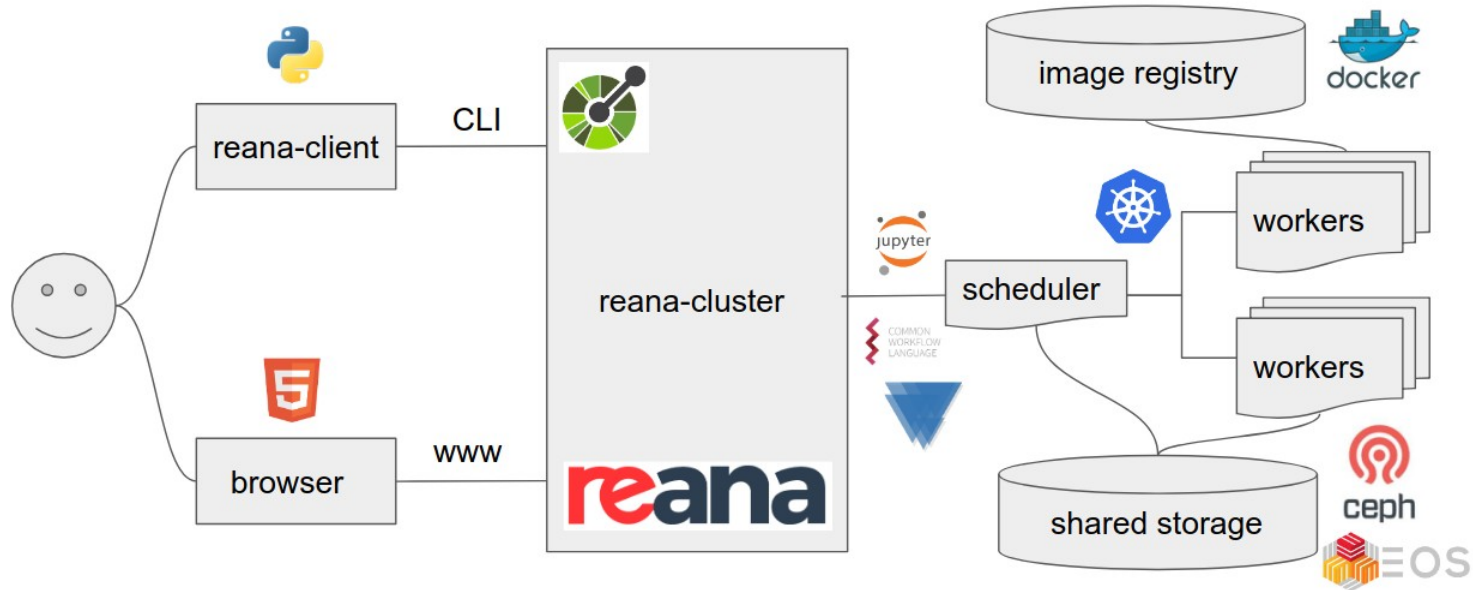
- CI/CD
 - Automatic run by changes
 - In this case → build docker image
- Main Idea
 - One large project defines all deployments
 - *.yaml file for rancher / fleet
 - Only trigger sub jobs

```
variables:
  IMAGE_TAG: $CI_REGISTRY_IMAGE/aipkube.legal:
  $CI_COMMIT_REF_NAME

build-push-docker-image-job:
  image: docker:20.10.16
  services:
    - docker:20.10.16-dind
  before_script:
    - docker info
  script:
    - cd aipkube/legal
    - export
    - ls -latr
    - echo $CI_REGISTRY_PASSWORD | docker login -u
      $CI_REGISTRY_USER $CI_REGISTRY --password-stdin
    - docker build -t $IMAGE_TAG .
    - docker push $IMAGE_TAG
```



3. REANA



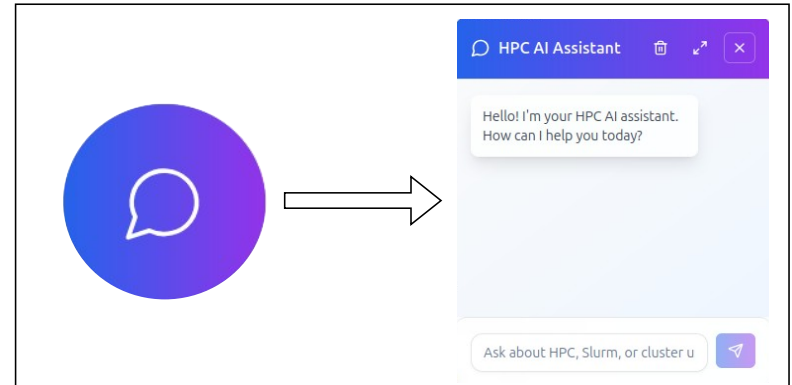
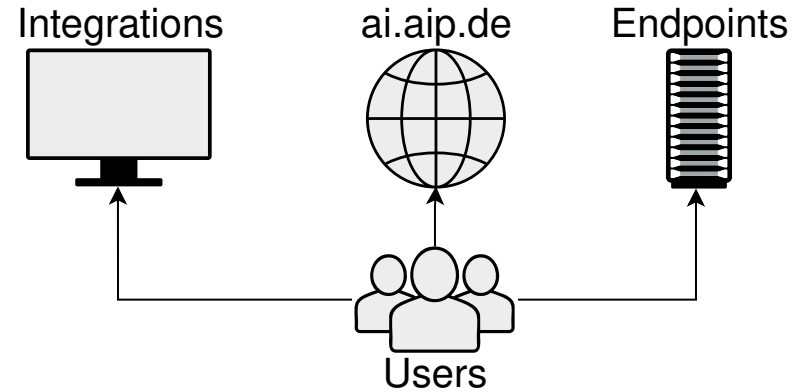
<https://docs.reana.io/images/reana-platform-20181202.png>

4. AI Goals / Purposes

- Develop an AI ecosystem for scientific purposes
 - Local and open source
- Purposes and ideas
 - Text generation, vibe coding
 - Agentic systems → generating complete data workflows / pipelines
 - Direct integrations → e. g. on websites

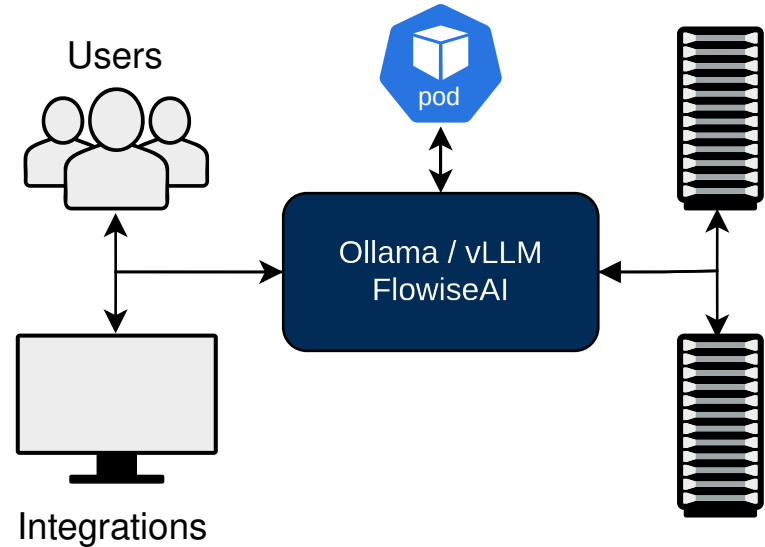
5. Users view

- ai.aip.de
 - OpenAI like chat interface via “Open WebUI”
- Endpoints
 - Multiple API endpoints
- Direct integrations
 - Like chat bots on websites



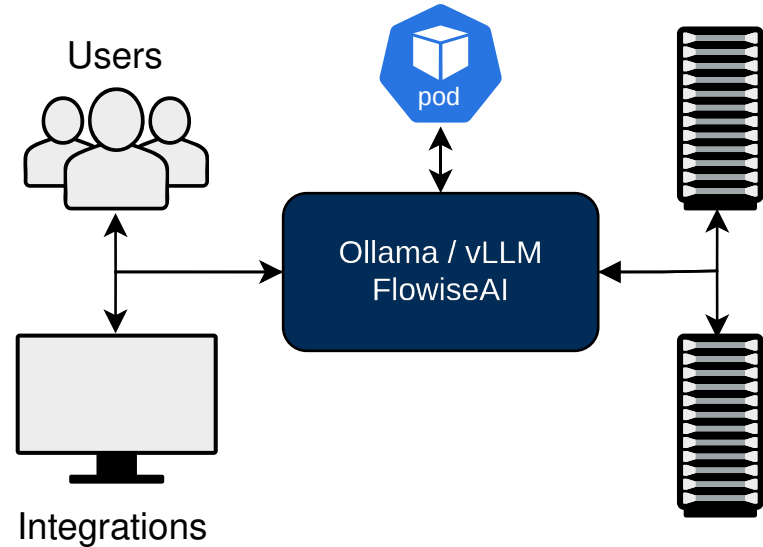
6. Integrations / Development

- AI assistant to generate REANA workflows
 - FlowiseAI as backend
 - Generates reana.yaml files
 - And code
 - Edit given workflows
 - Create GitLab repositories
- Integration
 - Connected to Open WebUI
 - As chat bot in „astro card hub“ (future)



6. Integrations / Development

- AI assistant to answer questions about HPC in AIP environments
 - FlowiseAI as backend
- Integration
 - As chat bot in „hpc.kube.aip.de“ (in development)



[← Back to articles](#)

Cluster Documentation

Intermediate

Newton Cluster User Guide

by HPC Team

July 8, 2025

6 min read

newton

cluster

slurm

gpu

hpc

Newton Cluster User Guide

Welcome to the comprehensive guide for using the **Newton 2021** cluster. This high-performance computing system uses **Slurm** (Simple Linux Utility for Resource Management) for job scheduling and resource allocation.

Overview

The Newton cluster is designed to handle computationally intensive tasks across multiple specialized node types, each optimized for different workloads.

Cluster Architecture

Node Categories

The Newton cluster consists of four distinct node categories, each tailored for specific

HPC AI Assistant

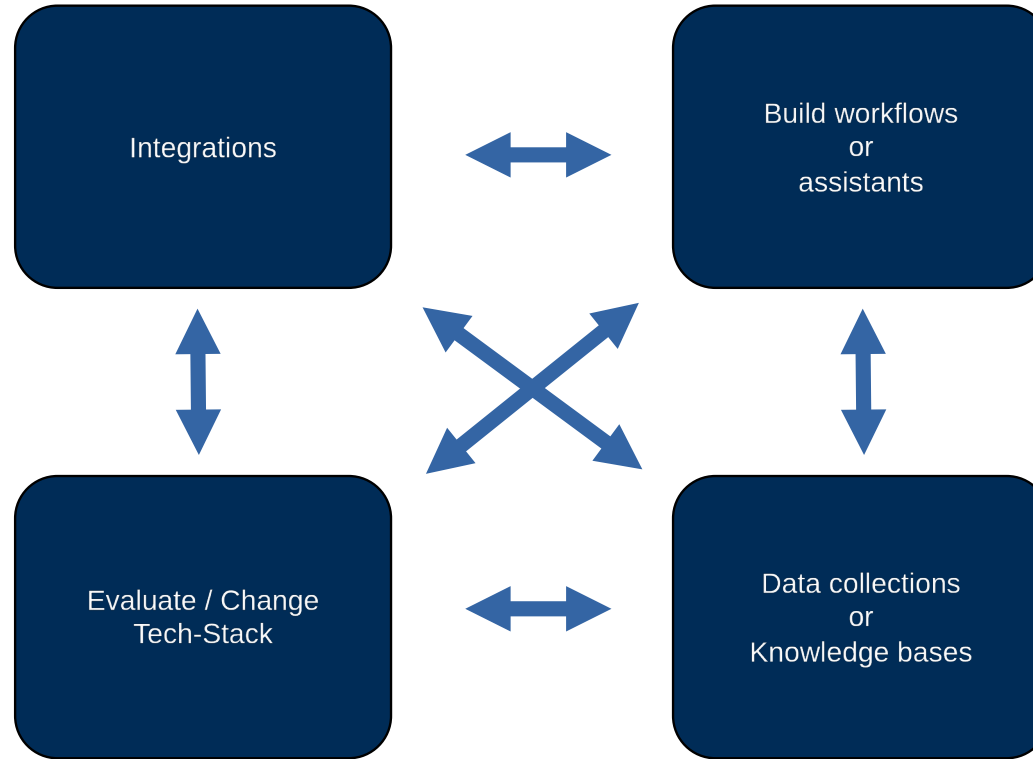
If you encounter any conflicts or issues with existing modules, you may want to purge the current environment first using `module purge` before loading new ones:

```
module purge
module load <module_name>
```

This ensures that your environment is clean and only contains the necessary modules.

Ask about HPC, Slurm, or cluster u

7. Outlook



8. Summary

Achievements

- Local hosted approach
 - Full control
 - GDPR
- Assistants
- Integrations

Problems / missing parts

- No reliable user data
 - Efficiency
 - Usage / metrics
- Data collection(s)
- GDPR
- User management
- Guides