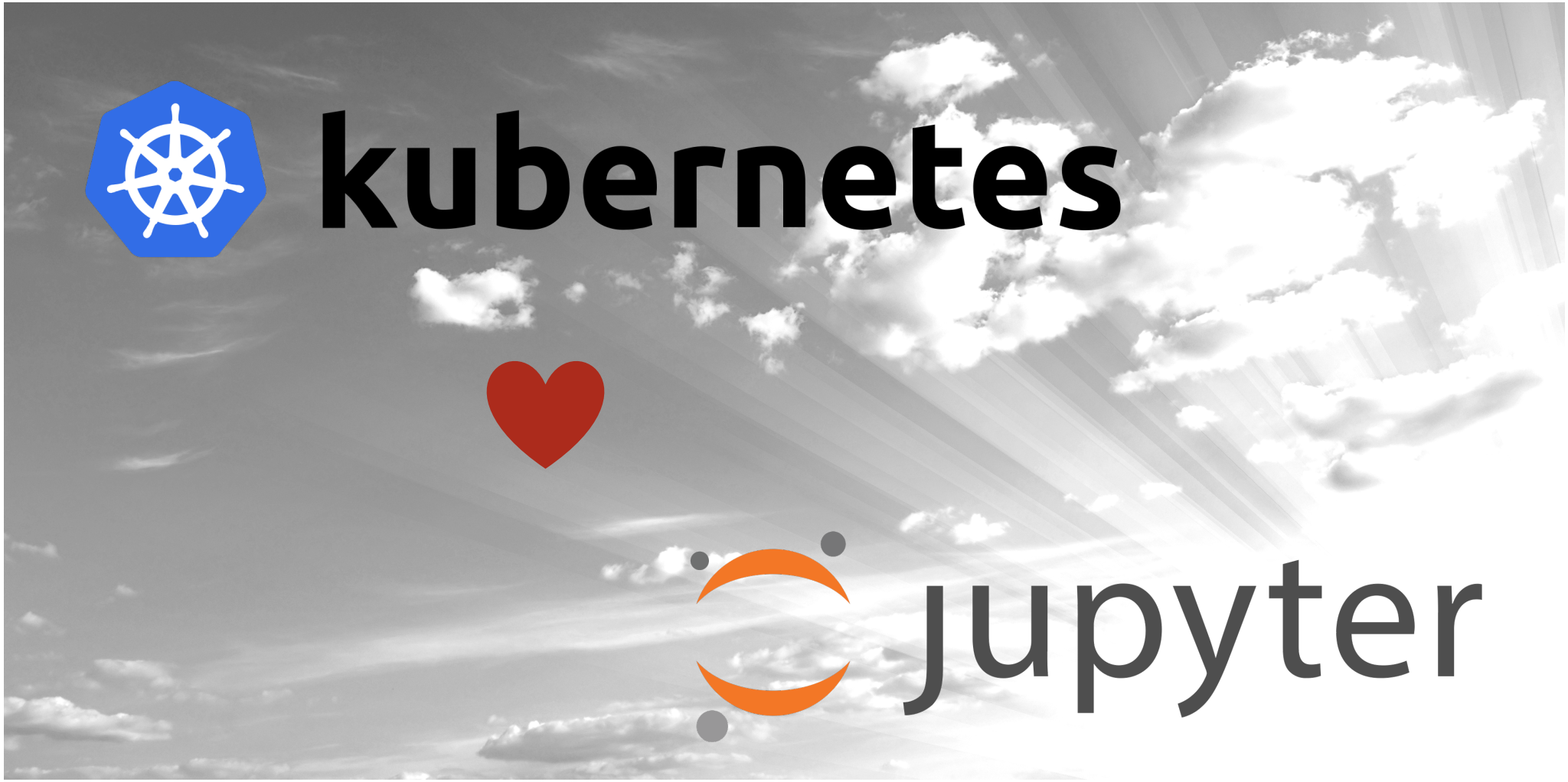


Lehren in der Cloud – Über den Aufbau einer eigenen Cloud-Infrastruktur

Steffen Vogel
svogel2@eonerc.rwth-aachen.de

ACS | Automation of Complex Power Systems
E.ON Energy Research Center
RWTH AACHEN UNIVERSITY

Automation of Complex Systems…

Inhalt

- Wie alles begann
  - OSByExample.com
- RWTH JupyterHub Cluster
  - Microservices
  - Statistiken
- Kubernetes
  - Netzwerk
  - Speicher
  - Management, Deployment, Monitoring
- **Lessons Learned**

OSByExample.com

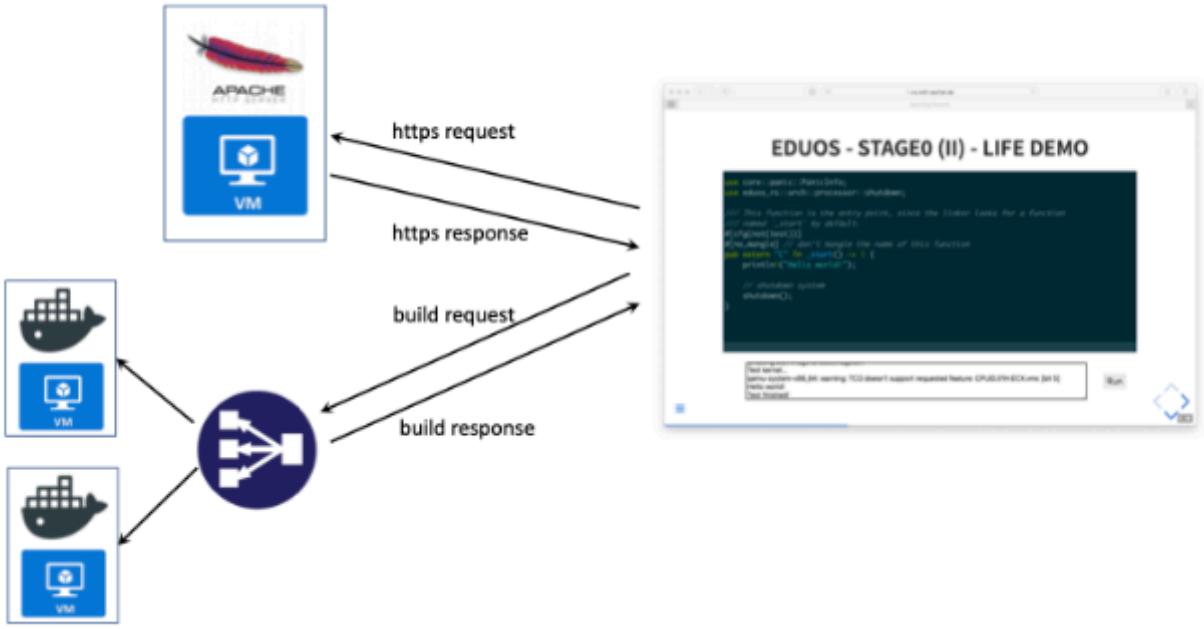
- Vorlesungsfolien als Website
  - https://slides.osbyexample.com

- Integration von interaktivem C und Rust Codefragmenten
  - Ähnlich wie Rust Playground
  - Zusätzlich Ausführung von eduOS-rs

- Frontend
  - ACE Javascript Editor
  - Asciidoctor & Reveal.js

- Backend
  - In Go geschrieben
  - Rust/C Compiler
  - Sandboxing

- Kommunikation über Web-Sockets



jupyter.rwth-aachen.de

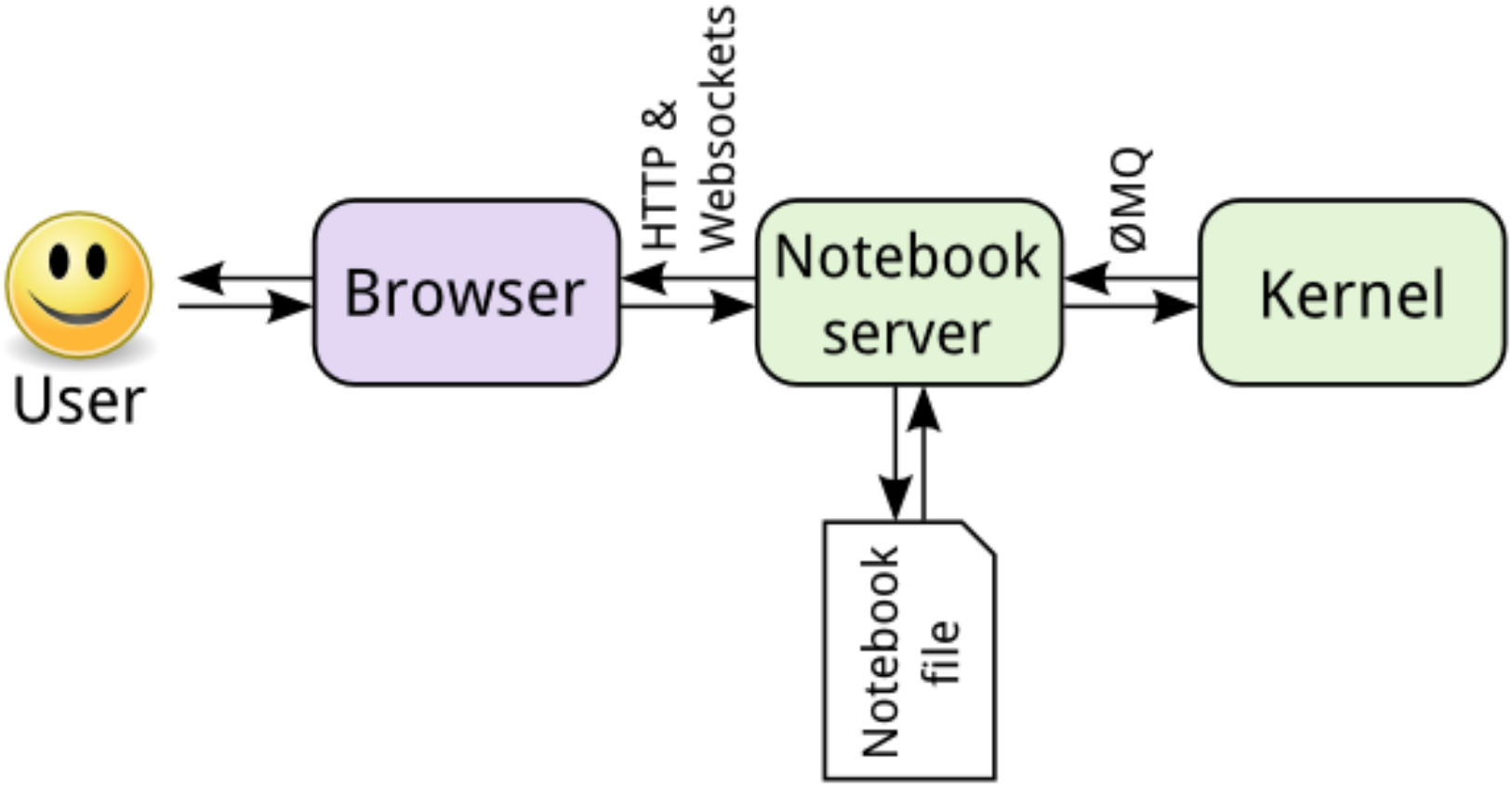
Ein JupyterHub Cluster für die RWTH

Jupyter Projekt: Geschichte

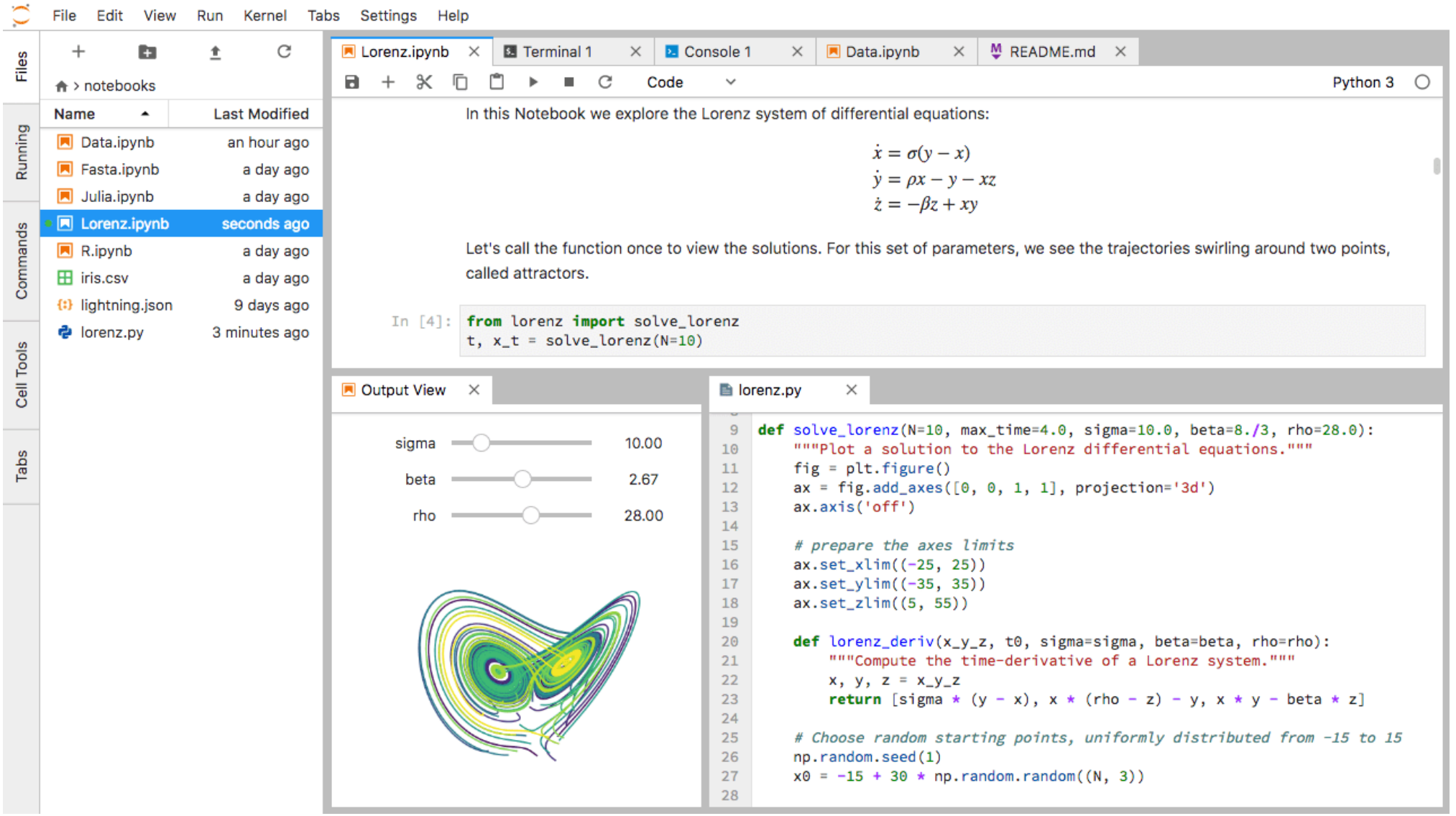
- IPython, 2001, Fernando Pérez
  - Interaktiver Python Interpreter
- IPython kernel (2010)
- IPython Notebook (prototype 2010/11, initial dev. 2005/7)

- **Jupyter Project** ab 2014
  - multi-language, beyond **Ju**lia, **Py**thon and **R** → „Jupyter“
    - = Mehr als 40 Kernel/Sprachen werden unterstützt
  - Non-profit, open-source, free to use
  - Jupyter Notebook App (server-client app)
  - Jupyter Hub (user managment & orchestration)
  - Next generation: Jupyter Lab
    - = New frontend for Notebook App

Jupyter Architecture



JupyterLab

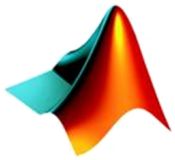


E.ON Energy Research Center | RWTH AACHEN UNIVERSITY

Unterstütze Jupyter Umgebungen (Kernel)

- Python
- OpenModelica
- Tensorflow
- pySpark
- Xeus Cling: C/C++
- R
- Simulation Elektrischer Netze: DPsim

- Möglicherweise:
  - MATLAB?
  - Lizenzen…

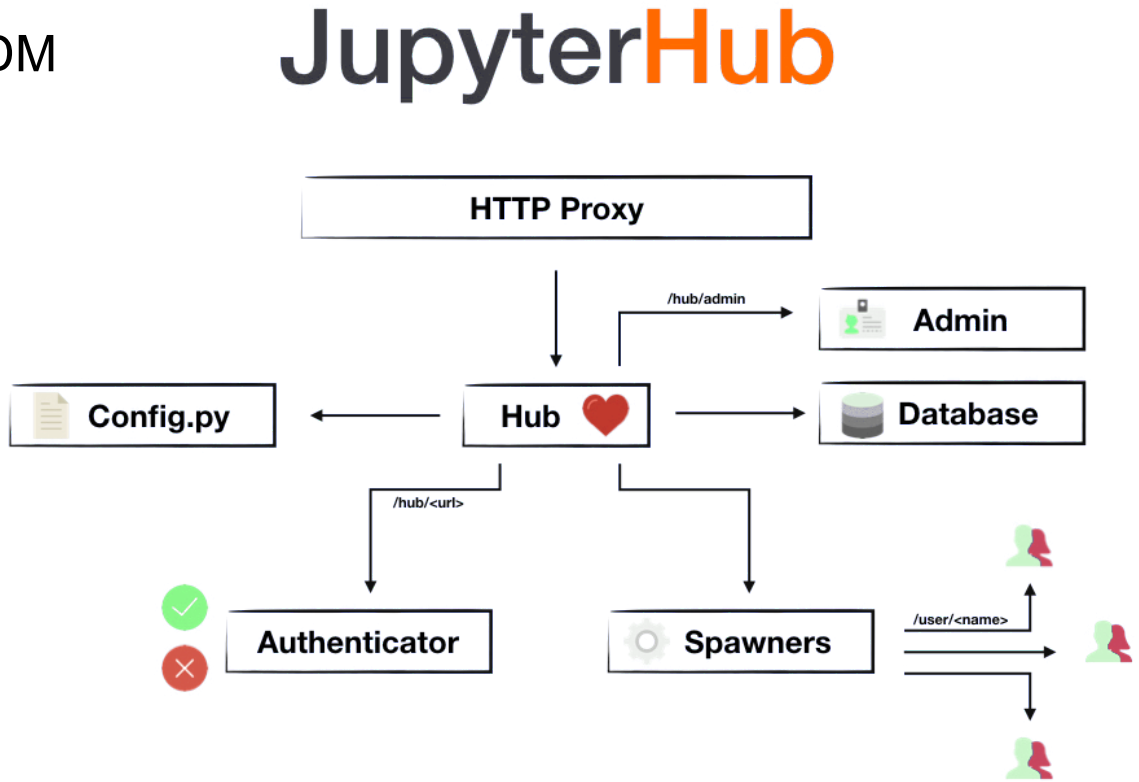


JupyterHub

- Mehr Benutzer Zugriff auf Jupyter Notebooks
- Startet und Verwaltet Singleuser Jupyter Container

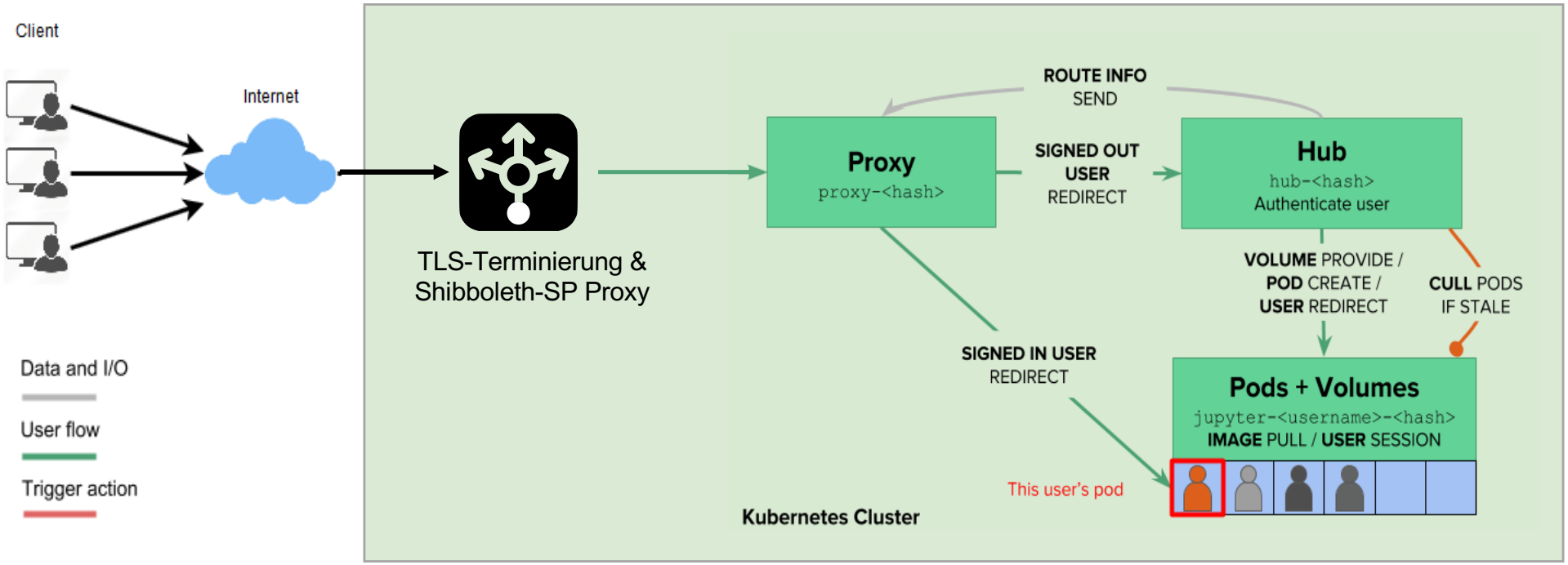
- Sehr gut erweiterbar

- Authentifizierung über RWTH IDM
  - Shibboleth



Zero 2 JupyterHub (Z2JH)

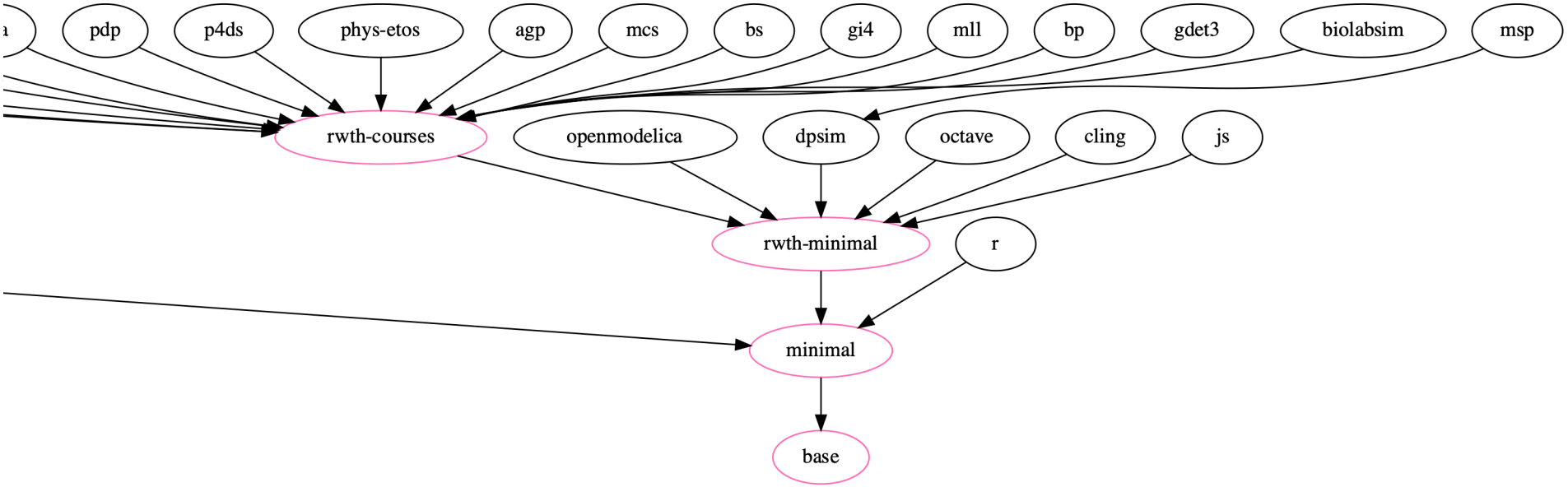
- *Zero to JupyterHub with Kubernetes*
  - JupyterLab Instanzen warden als K8S Pods verwaltet
  - Autoskalierung
  - Sandboxing
  - https://z2jh.jupyter.org



Profile

- Lehrveranstaltungen & Praktika definieren ihre Umgebung über Docker Images
- Dozenten liefern Git-Repo mit Dockerfile und Jupyter Notebooks
- Profil Metadaten werden zentral durch YAML Dateien verwaltet

```yaml
---
slug: mcs
display_name: "[MCS] Modern Control Systems"
priority: 55
```

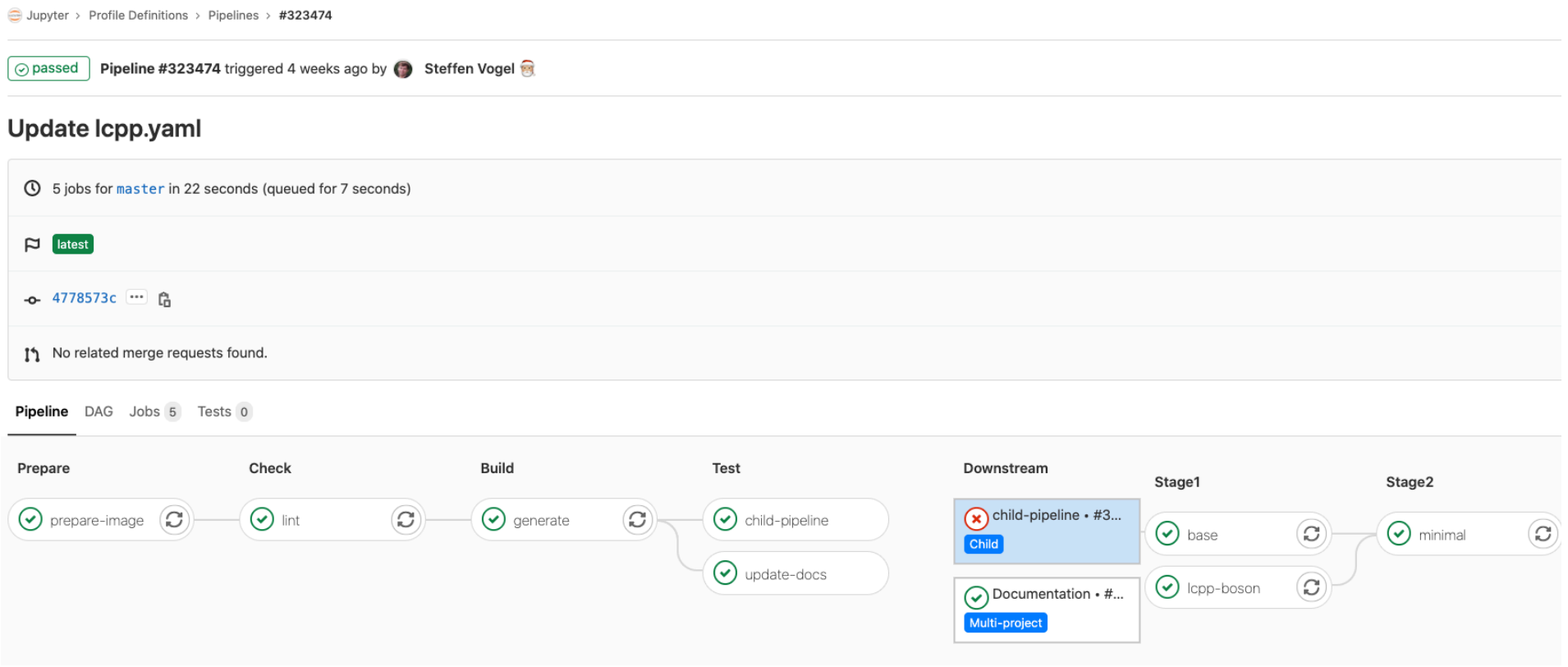



```yaml
base: rwth-courses
repo:
  url: https://git.rwth-aachen.de/acs/public/teaching/mcs/students.git
dockerfile: singleuser-octave.Dockerfile
```

- https://git.rwth-aachen.de/jupyter/profiles

Continous Integration

- GitLab CI Runner bauen Docker Images
- Image layering
- Vereinfachtes Lifecycle Management von Profilen
  - Sicherheits Patches!



Zielgruppe und Nutzer

- RWTH
  - 9 Fakultäten
  - 45000 Studierende
  - 162 Studiengänge

- **SS2019:** Proof-of-Concept
  - 3 Vorlesungen
  - Eigenes OpenStack Cluster

- **SS2020:** Pilotbetrieb
  - Anschaffung neuer Hardware
  - Zusammenarbeit RWTH ITC Center
    - = Erstes Produktiv Kubernetes Cluster
  - Mit ca. 20 Vorlesungen und Praktika

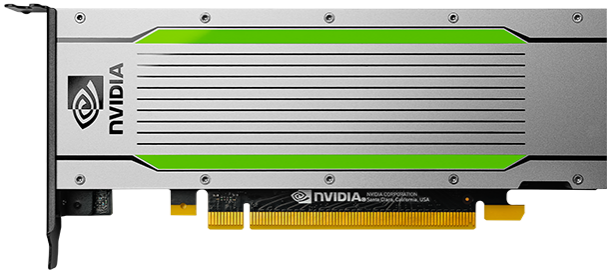
RWTH Kubernetes Cluster
Hardware

- Ziel: Viele Nutzer bediehnen können
  - Niedriger CPU Takt, Viele Cores

- 6 Knoten: Dell PowerEdge 740XD
  - Zwei Sockel Systeme:
    - 2x 16C / 32T Xeon Gold 5218 2,3 Ghz
  - 768 GB DDR4 RAM / node (5.376 GB total)
  - 5x 3.84 TB SSD Storage / node (134 TB total)

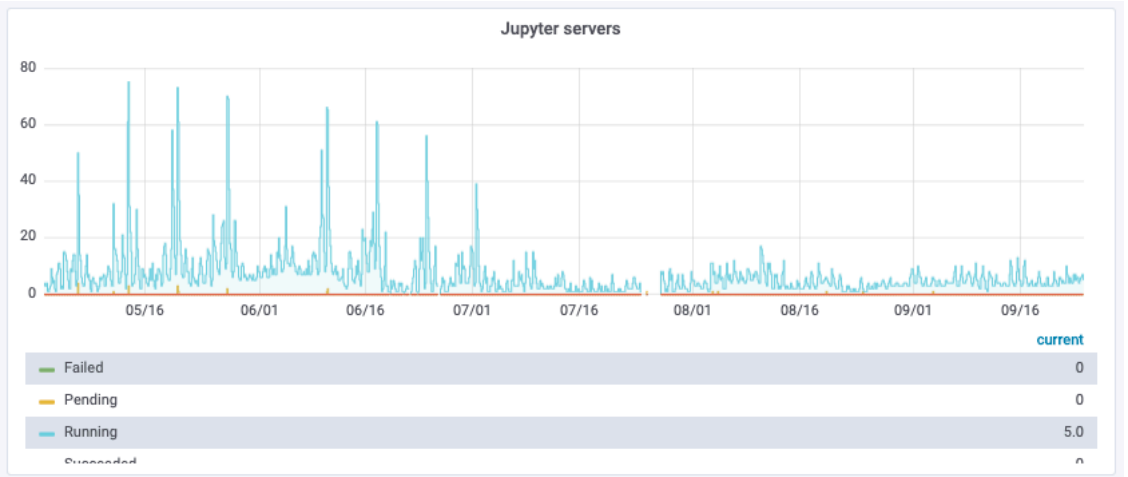
  - Redundant Dual 10 GigE links
    - LACP Link Aggregation nach IEEE 802.3ad
    - Virtual Gateway (Cisco VRRP / ACI)

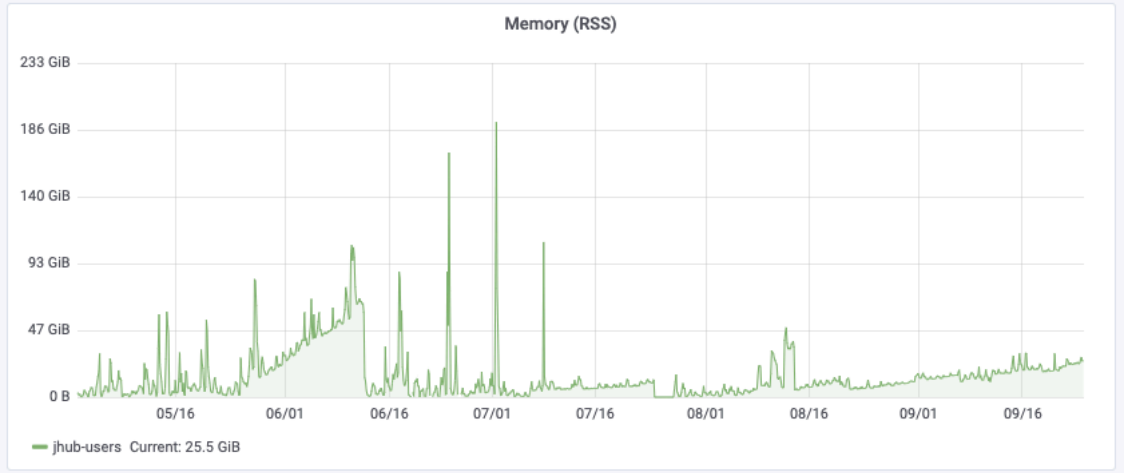
- 1 GPU Knoten
  - Wie oben, nur mit zusätzlich
  - 2x NVIDIA Tesla T4 GPGPUs



Statistiken

- Zeitraum: Mai – Sept 2020 (Pilotphase)
- Nutzer: 655
- Spawns: > 4400
- Max. gleichzeitige Nutzer: 76
- Profile:
  - Vorlesungen: 15
  - Praktika: 5
  - Generisch: 13





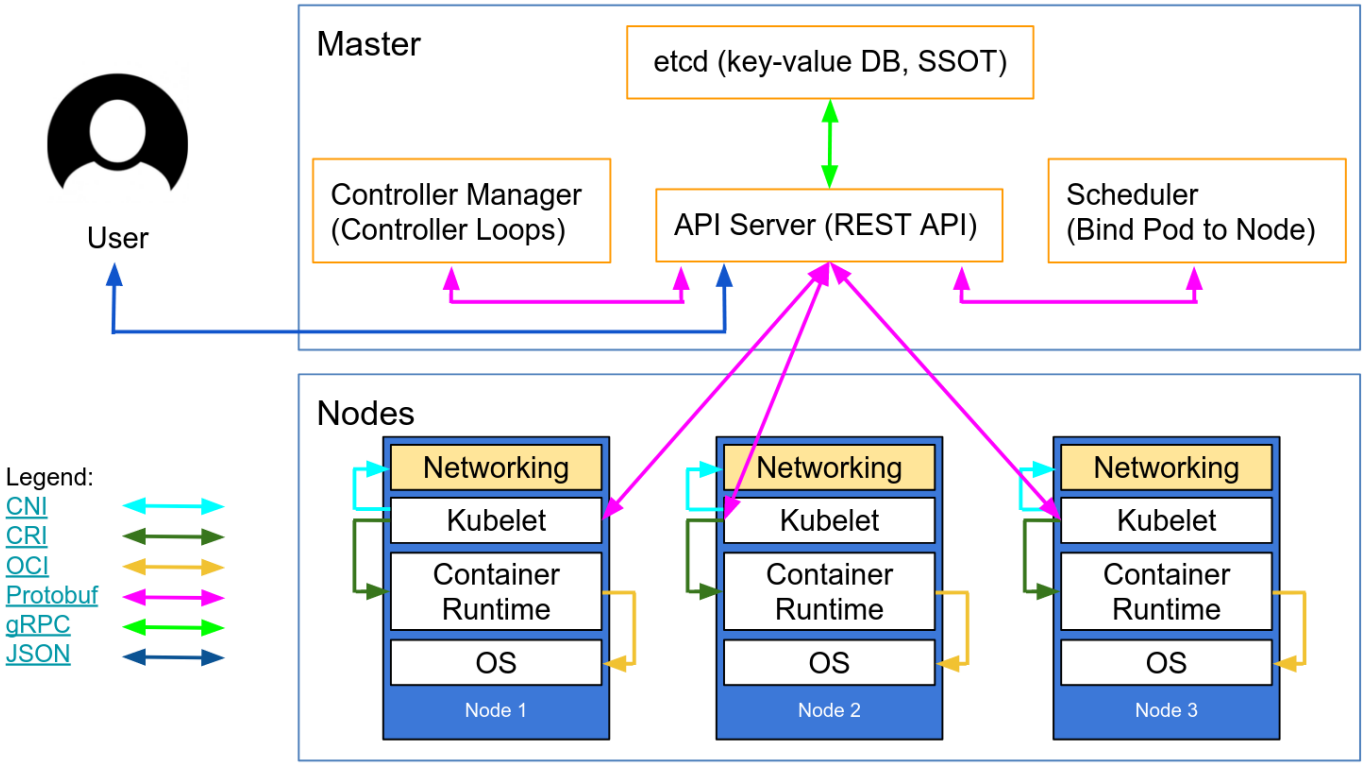
- https://grafana.jupyter.rwth-aachen.de/

Kubernetes

Kubernetes

- Ein Thema für sich…
- JupyterHub wird das erste Production K8S Cluster am ITC der RWTH 🥳

- Komplexität explodiert:
  - über 130 Pods für Orchestrierung, Monitoring, Storage, Network, …



Netzwerk

- Project Calico
  - Linux IPVS oder eBGP für Load Balancing

- Kubernetes Cluster ist IPv6 only
  - Ausschließlich Ingress Knoten haben eine IPv4 Adresse
  - Globale IPv6 Addressierung für
    - Pods / Containers
    - Services
  - Ermöglicht Nachverfolgbarkeit von Nutzern <-> Container Ips

- DNS64 & NAT64 für Zugriff auf IPv4 Endpunkte
- Last Verteilung / Hochverfügbarkeit
  - Layer 3: DNS Round Robin
  - Layer 4: Kubernetes Services

Speicher

- Hyperkonvergente Infrastruktur

- Ceph Object Storage
- Orchestrierung via Rook.io
  - Alle Ceph Storage Daemons laufen als Workloads im Kubernetes

- Persistente Benutzer Verzeichnisse
  - Ceph RBD Block Devices
  - Replizierter Pool mit jeweils 3 Replika pro Block
  - ReadWriteOnce (RWO)
  - → Wir profitieren vom Linux Pagecache

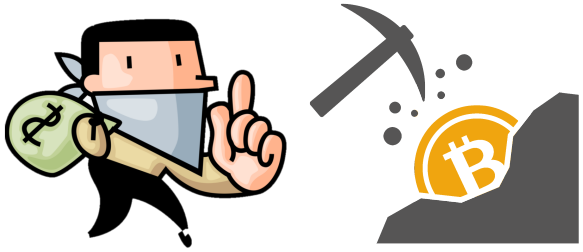
- Geteilte Verzeichnisse
  - Datasets
  - Beispiele

Kubernetes
Management + Provisioning

- Baremetal Provisioning
  - Puppetlabs Razor
  - DHCP + PXE Boot
- Deployment
  - Ansible Playbooks
  - Kubeadm
- Management
  - Helm
  - Rancher
- Monitoring
  - Prometheus
  - Grafana
  - Logging: Elasticsearch, Fluentd, Kibana
- https://git.rwth-aachen.de/jupyter/provisioning

Lessons Learned: Jupyter

- Jupyter kann weit mehr als nur Python
  - Cling C/C++ Kernel
  - Web Terminal & Editor
  - Eigene Panels, Widgets, …
- Missbrauch
  - Crypto Miner
    - Bisher hatten wir Glück
    - Heben sich gut von der gewöhnlichen Nutzung ab.
  - Filesharing?
- Quotas...
  - Gurobi explodiert gerne mal..



Lessons Learned:
Effiziente Speicher Ausnutzung

- Gleichartige Workloads
  - Hunderte - Tausende **X** Interpreter (**X** ∈ {Python, Octave, Julia, R})
  - Praktika: bis zu Hunderte Studierenden arbeiten auf gleichen Datensätzen

Frage: Funktionierende existierende Methoden zur Effizienzsteigerung der Speicher Nutzung auch in Container Umgebungen?

- Memory Mapped Files
  - Dynamische Lader (/lib/ld.so) nutzt mmap()
  - Copy-on-Write
  - Python Interpreter sowie Module werden als Dynamische Bibliotheken geladen
  - mmap()
    - = lazy-loading
    - = copy-on-write

- Kernel Samepage Merging (KSM)
  - Deduplizierung von Speicherseiten
  - Ursprünglich für VMs entwickelt
  - Möglicherweise sinnvoll für Deduplizierung von geladenen Working Sets?
  - KSM Daemon, Periodische Scans

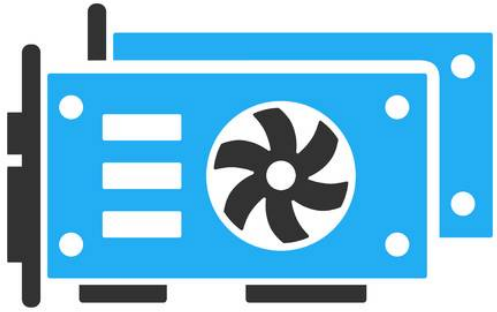
- Wie sieht es auf GPUs aus?
  - GPU Samepage Merging (GSM)?

Lessons Learned:
Kubernetes GPU Share Scheduler Extender

- Machine Learning / AI Workloads
- 1 GPU pro Container ist unrealistisch

- Ansätze
  - Nicht-interaktive Batch-verarbeitung von Notebooks
    - = SLURM, Dask, …
  - Mehr GPUs ($$$)
  - Resource Sharing, Overcommit?

- Alibaba Cloud hat Kubernetes Scheduler erweitert
  - Bisher keine Speicher Isolation
  - **Lösung:** NVIDIA Multi-Process Service (MPS)
    - = Co-operatives Multi-tasking auf der GPU



Lessons Learned: Skalierbarkeit

- JupyterHub hält lokalen Zustand vor
  - Skaliert leider nicht
  - Mehrere Replika nicht möglich
  - Rolling Updates nicht möglich
  - **Lösung:** Hub Pod muss Zustandslos werden
    - Redis KV Store als schnellerer Cache

- Reverse Proxy (CHP) skaliert noch nicht
  - **Lösung:** Alternative Proxies
    - Traefik
    - Kubernetes Ingress Controller

Links

- https://jupyter.rwth-aachen.de
  - https://jupyter.pages.rwth-aachen.de/documentation
  - https://git.rwth-aachen.de/jupyter
- https://osbyexample.com
- https://github.com/AliyunContainerService/gpushare-scheduler-extender
  - https://www.alibabacloud.com/blog/gpu-sharing-scheduler-extender-now-supports-fine-grained-kubernetes-clusters_594926
- https://kubernetes.org
  - https://ceph.com
  - https://rook.io
  - https://www.projectcalico.org
- https://jupyter.org
- https://zero-to-jupyterhub.readthedocs.io

Contact

E.ON Energy Research Center
Mathieustraße 10
52074 Aachen
Germany

Steffen Vogel
T +49 241 80 49577
F +49 241 80 49709
svogel2@eonerc.rwth-aachen.de
http://www.eonerc.rwth-aachen.de

ACS | Automation of Complex Power Systems | E.ON Energy Research Center | RWTH AACHEN UNIVERSITY