

XAAS *FROM IDEAS
TO IMPACT.*

XAAS AFTERHOURS

25.06.2026



ABLAUF

1 *Daniel Menzel: OpenStack für HPC-Workloads*

2 *Ilja Livenson: Orchestrating HPC Allocations with Waldur*

3 *Panel: Two Sides of the same Cloud*

4 *Network & Grillen*

PANEL – TWO SIDES OF THE SAME CLOUD

EINE DISKUSSION MIT:



ILJA LIVENSON



DANIEL MENZEL



ALEXANDRA HEINRICH

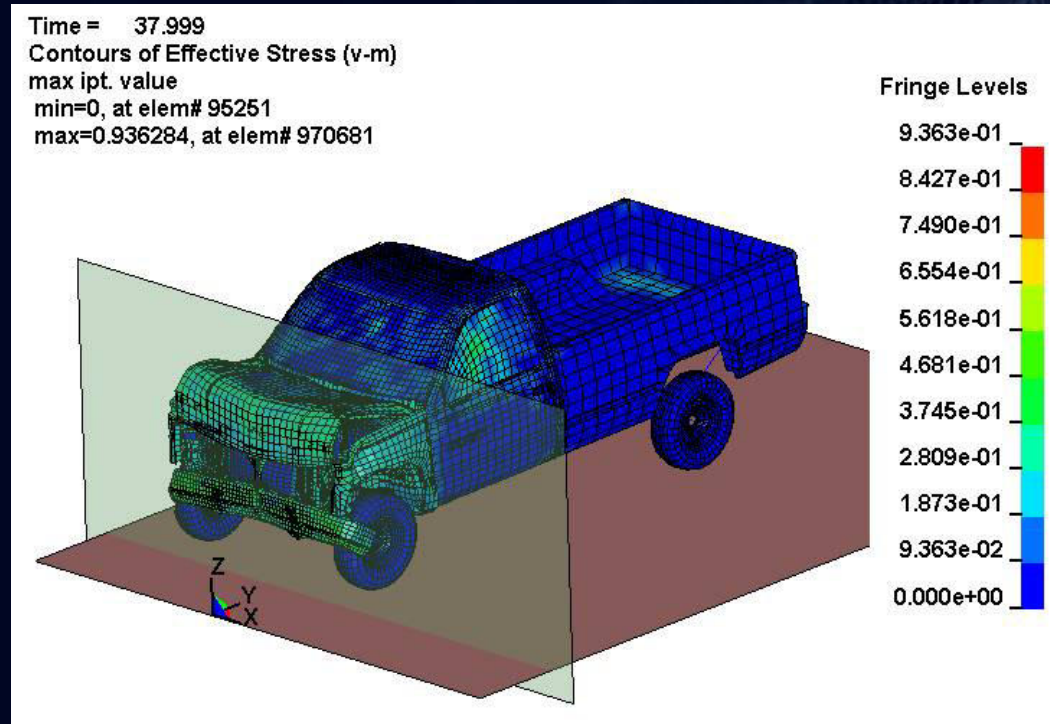
OpenStack

für High-Performance-Workloads



XAAS

„high performance workload“?



„high performance workload“?

$$\sum_{i=1}^{12} i = 1+2+3+4+5+6+7+8+9+10+11+12$$

„high performance workload“?

RESOURCES?



EVERYTHING

„high performance workload“?

 deepseek-v4-flash-max ▾ +



Ich halte heute einen Vortrag zum Thema "OpenStack für hochperformante Workloads" in Hamburg. Bitte begrüße das Publikum mit einem freundlichen Moin und erzähle ihm einen humorvollen aber korrekten "random fact". Suche dafür im Internet, wenn nötig. Duze uns.

 **deepseek-v4-flash-max** Heute um 12:21

 Untersucht: 17 search_web, 2 fetch_url ▾

Moin zusammen!

Schön, dass ihr alle heute nach Hamburg gekommen seid, um über „OpenStack für hochperformante Workloads“ zu sprechen. Bevor wir richtig loslegen, habe ich einen kleinen Fun Fact für euch – der ist nicht nur unterhaltsam, sondern auch 100 % korrekt (versprochen!).

Ihr wisst ja vielleicht, dass OpenStack-Releases traditionell nach Orten benannt werden, in denen der jeweilige Design Summit stattfand – angefangen mit **Austin** (Texas), dann **Bexar**, **Cactus**, **Diablo** ... alles schön alphabetisch und geografisch. Aber dann kam die berühmte „**Waldon Exception**“ [Release Naming ...Stack wiki +1](#) : Weil der Grizzlybär auf der kalifornischen Flagge einfach zu cool klang, durfte plötzlich auch ein nicht-örtlicher Name gewählt werden – und schwupps, hieß ein Release „**Grizzly**“.

Noch besser wurde es beim „**Kilo**“-Release [Release Naming ...Stack wiki +1](#) : Der Name kommt nicht etwa von einer Stadt, sondern vom **offiziellen Referenz-Kilogramm** in Sèvres bei Paris – jenem Zylinder aus Platin-Iridium, der bis 2019 das letzte physikalische Artefakt war, das eine SI-Basiseinheit definierte. Ein OpenStack-Release, benannt nach einem Klotz in einem Tresor in Frankreich. Wenn das kein IT-Nerd-Humor ist, weiß ich auch nicht! 😊

Also: OpenStack kann nicht nur hochperformante Workloads, sondern auch extravagante Namensgebung.

Lasst uns eintauchen!

 67 Quellen





Eine Nachricht senden

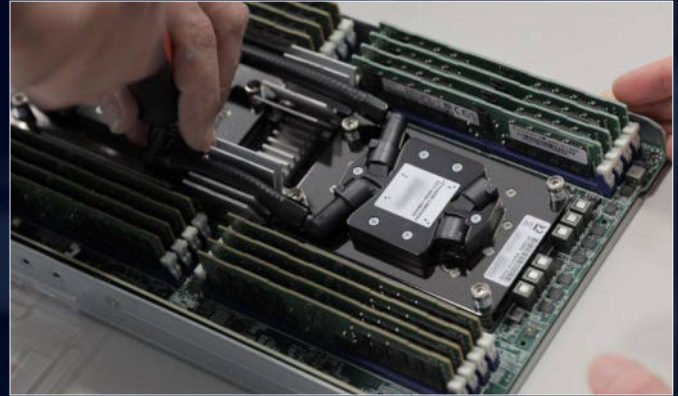
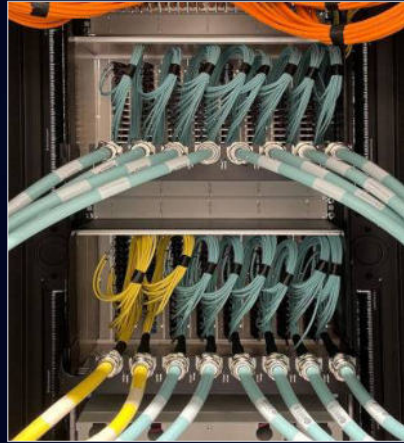
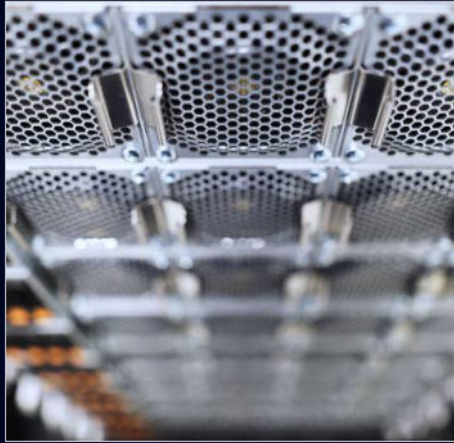




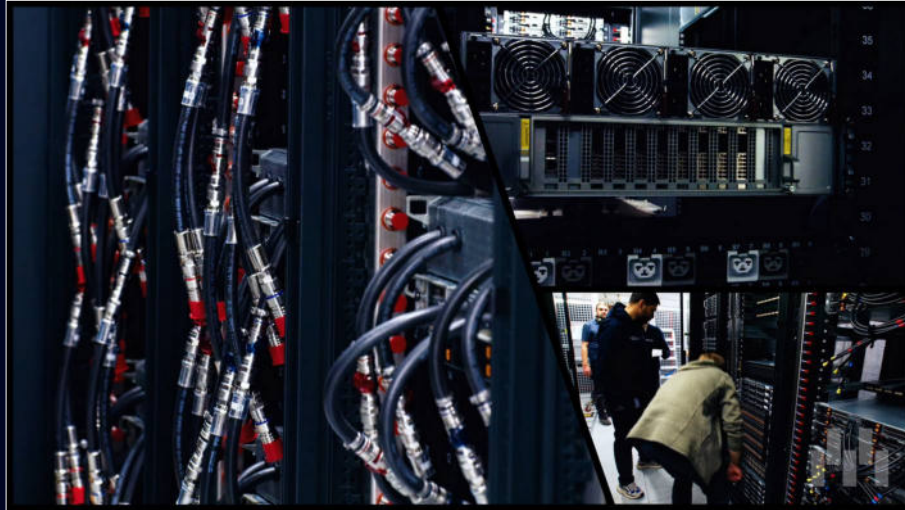
„high performance workload“?



Heute HPC & AI, morgen Enterprise-IT



Training vs Inferenz



- NVIDIA B200
- SXM/NVLink
- 4.500 TFLOPS/GPU
- 180 GB VRAM/GPU
- 1.200 W/GPU
- 10x 400G
- ~40 T€/GPU



- RTX6000 Blackwell (Max-Q)
- PCIe
- 125 TFLOPS/GPU
- 96 GB VRAM/GPU
- 300 W/GPU
- 1x Dual 25G
- ~9 T€/GPU

HPC-Paradigmen

- Standardisierung – klassisch ein Monolith



HPC-Paradigmen

- Standardisierung – klassisch ein Monolith
- Alles, was Performance bringt: 🥰
 - RDMA + 800G
 - moderne CPUs
 - All-Flash-Systeme
- Alles, was Performance kostet: 😡
 - Authentifikation
 - Encryption

Cloudparadigmen

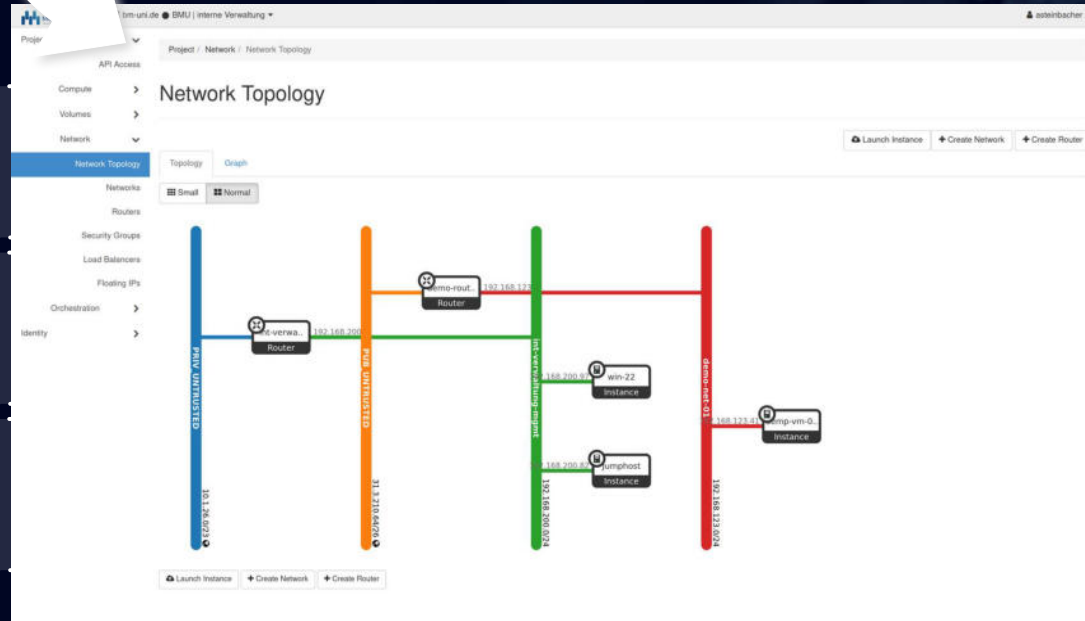
- Standardisierung = cool
- ein Tenant darf den anderen **nicht beeinflussen**
- Self-Service, wo immer es geht
- für den User ist die Cloud **einfach zu benutzen**
- → alle Funktionen **virtualisiert in der Cloud**

Overlay (Workload)

CPU-VMs

GPU-VMs

Netzwerke



LBaaS

DNSaaS

VPNaaS

Server

Switches

Storage

Strom

Underlay

zusätzliche Services

Loadbalancer as a Service (Octavia)

Menzel IT GmbH

mgpv.de

fclauss-testing

fclauss

Project

API Access

Compute

Volumes

Network

Network Topology

Networks

Routers

Security Groups

Load Balancers

Floating IPs

Orchestration

Identity

Project / Network / Load Balancers

Load Balancers

Q

Click here for filters or full text search.

x

+ Create Load Balancer

Delete Load Balancers

Displaying 2 items

☐	Name ^	IP Address	Availability Zone	Operating Status	Provisioning Status	Admin State Up	
☐	▼ ha-lb	10.0.5.171	-	Online	Active	Yes	Edit Load Balancer
Name ha-lb		Created At 2025-10-12T13:14:00		Network ID d363ef3d-7bb9-4232-8e45-eaf6902dde17		Flavor ID -	
ID e11c5867-eecc-4cc7-9c24-a06f0ef80477		Updated At 2025-10-12T14:31:03		Subnet ID b59a084e-8aef-4f54-94cc-beee9c4f4d00		Provider amphora	
Project ID 7befb8c4b76f4bcd92be5c6ed52d2497		Description None		Port ID 324a8ed2-f88e-486d-8a39-b3ac0eed73cc		Floating IP 10.1.27.2	
☐	► Web-Loadbalancer	10.0.5.32	-	Online	Active	Yes	Edit Load Balancer

Displaying 2 items

zusätzliche Services

Router

Menzel IT GmbH | bm-uni.de | BMU | interne Verwaltung | asteinbacher

Project / Network / Routers

Routers

Router Name = Filter [+ Create Router](#) [Delete Routers](#)

Displaying 2 items

☐	Name	Status	External Network	Admin State	Availability Zones	Actions
☐	demo-router-01	Active	PUB_UNTRUSTED	UP	-	Clear Gateway
☐	int-verwaltung-router	Active	PRIV_UNTRUSTED	UP	-	Clear Gateway

Displaying 2 items

zusätzliche Services

Floating IPs (externe IP-Adressen an Ressourcen binden)

Menzel IT GmbH

bm-uni.de

BMU | interne Verwaltung

asteinbacher

Project

API Access

Compute

Volumes

Network

Network Topology

Networks

Routers

Security Groups

Load Balancers

Floating IPs

Orchestration

Identity

Project / Network / Floating IPs

Floating IPs

Floating IP Address = Filter [Allocate IP To Project](#) [Release Floating IPs](#)

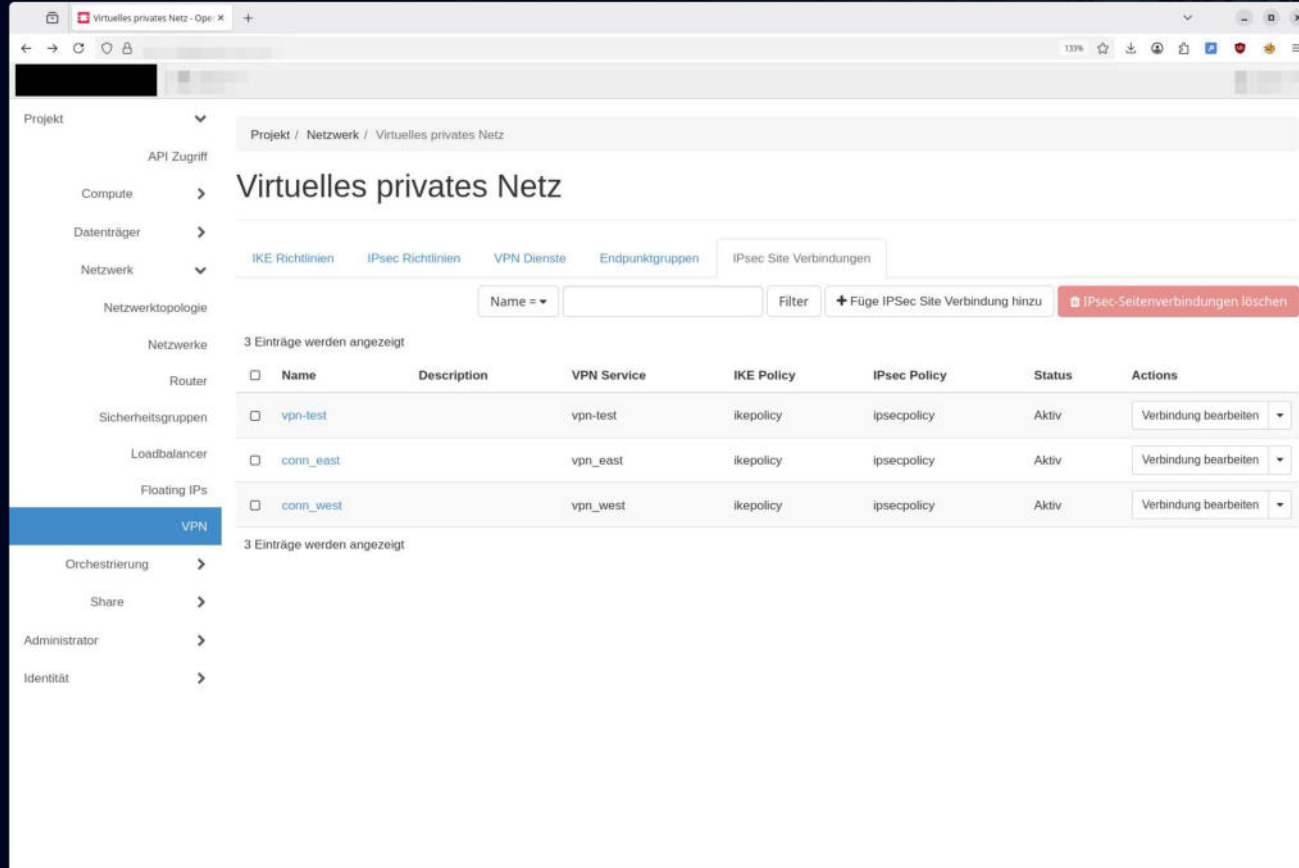
Displaying 2 items

☐	IP Address	Description	DNS Name	DNS Domain	Mapped Fixed IP Address	Pool	Status	Actions
☐	10.1.27.25				Jumphost 192.168.200.82	PRIV_UNTRUSTED	Active	Disassociate
☐	31.3.210.122				demp-vm-01 192.168.123.41	PUB_UNTRUSTED	Active	Disassociate

Displaying 2 items

zusätzliche Services

VPN-as-a-Service



The screenshot shows the VMware vSphere Web Client interface for configuring a Virtual Private Network (VPN). The breadcrumb navigation indicates the path: Projekt / Netzwerk / Virtuelles privates Netz. The main title is 'Virtuelles privates Netz'. Below the title, there are tabs for 'IKE Richtlinien', 'IPsec Richtlinien', 'VPN Dienste', 'Endpunktgruppen', and 'IPsec Site Verbindungen'. The 'VPN Dienste' tab is currently selected. A search bar with 'Name =' and a 'Filter' button is present. A red button 'Füge IPsec Site Verbindung hinzu' and a red button 'IPsec-Seitenverbindungen löschen' are also visible. The table below shows 3 entries displayed, with columns for Name, Description, VPN Service, IKE Policy, IPsec Policy, Status, and Actions.

☐	Name	Description	VPN Service	IKE Policy	IPsec Policy	Status	Actions
☐	vpn-test		vpn-test	ikepolicy	ipsecpolicy	Aktiv	Verbindung bearbeiten
☐	conn_east		vpn_east	ikepolicy	ipsecpolicy	Aktiv	Verbindung bearbeiten
☐	conn_west		vpn_west	ikepolicy	ipsecpolicy	Aktiv	Verbindung bearbeiten

Below the table, it states '3 Einträge werden angezeigt'.

zusätzliche Services

Fileshare-as-a-Service (Manila)

Create Share

Share Name *

demo-share

Beschreibung

Share Protocol *

NFS

Größe (GiB) *

20

Share Type *

default_share_type

Verfügbarkeitszone

nova

Share Group

Share Network *

demo-share

Metadaten

☐ Make visible to users from all projects

Beschreibung:

Select parameters of share you want to create.

Metadaten:

One line - one action. Empty strings will be ignored.
To add metadata use:

key=value

Share Limits

Gesamt Gibbytes 10 von 1.000 GiB verwendet

Number of Shares 1 von 50 verwendet

Abbrechen Erstellen

zusätzliche Services

GPU-Flavors

The screenshot displays the OpenStack Instanzkonsole interface. On the left, a sidebar lists various services: Projekt, API Zugriff, Compute, Übersicht, Instanzen (highlighted), Abbilder, Schlüsselpaare, Servergruppen, Datenträger, Netzwerk, Orchestrierung, and Identität. The main window shows the 'Instanzkonsole' for instance '000000ad'. A blue banner at the top provides instructions on how to interact with the console. Below this, a terminal window shows the connection to a QEMU instance. The primary view is a Blender 4.2.3 LTS window. The 'Blender Preferences' dialog is open, specifically the 'Cycles Render Devices' tab. Under 'Cycles Render Devices', 'CUDA' is selected, and 'NVIDIA A100-SXM4-40GB' is chosen. A red arrow points to this selection. The 'Operating System Settings' and 'Network' sections are also visible. In the background, a 3D model of an NVIDIA A100 GPU card is shown, with lines indicating its connection to the system. The Blender interface also shows the 'Scene' and 'Properties' panels on the right.

Tuning I

- schnellere Hardware
- Ionic
- Container
- NUMA-Binding
- SR-IOV
- DPUs (Bluefield)

Tuning II

- die 3 Ebenen
 - KVM kann das
 - OpenStack kann das
 - kolla-ansible kann das
(oder ein anderes *Deployment-Framework*)



Menzel IT GmbH

Charlottenburger Str. 33a
13086 Berlin

+49 30 5130 444 – 00

post@menzel-it.net

<https://menzel-it.net>



Waldur for sovereign AI on OpenStack

Turning sovereign OpenStack infrastructure into a governed, self-service platform for AI workloads.

Who is talking, and about what

OpenNode

Open-source company in Tartu, Estonia. We build and support Waldur.

Waldur

An open-source (MIT) cloud and HPC management platform with a self-service marketplace. Generic by design, sharpened for private clouds and research computing.



~10 yrs

in production



MIT

licensed, no lock-in



1 portal

many backends



EU-built

yours to run

WHY THIS, WHY NOW

Sovereign AI stopped being a slogan

Four pressures are pushing AI compute back onto infrastructure you actually control:



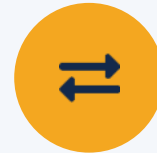
Data residency

GDPR, the EU AI Act and NIS2 make where data and models live a legal question, not a preference.



Public investment

EuroHPC AI Factories put national GPU capacity online — and it has to be shared on Europe's terms.



Lock-in fatigue

Organisations want an exit. Open stacks and portable workloads beat proprietary AI clouds.



Accountability

Who accessed which model and data, under what terms — auditable, not assumed.

THE REQUIREMENTS

What “sovereign” actually demands of the stack



Control over data

Workloads stay on infrastructure you own and operate.



An open, auditable stack

No black boxes you cannot inspect or replace.



Self-determined access

You decide who gets in, to what, and on which terms.



Traceability

Every allocation, consent and usage event is recorded.



No lock-in

Portable workloads and a credible exit at any time.



Cost transparency

Real usage and billing per project, team and tenant.

THE CORE IDEA

OpenStack is sovereign IaaS — not yet a platform



Users & organisations

researchers · enterprises · public sector



Waldur — orchestration & governance layer

onboarding · marketplace · quotas · identity · policy · billing



OpenStack

sovereign compute · storage · network · GPUs

The gap

OpenStack hands you tenants, flavours and volumes. It does not hand you a way to onboard users, approve and meter allocations, enforce policy, or bill — without tickets and tribal knowledge.

Waldur fills that gap and stays out of the data path: orchestration outside, policy inside.

THE PLATFORM

Waldur in one slide

A self-service marketplace and control plane that sits above your infrastructure.



Self-service marketplace

Catalogue of offerings users order from, with approval workflows.



Multi-tenant projects

Organisations, projects and roles map cleanly onto OpenStack tenants.



Metering & billing

Usage tracked and invoiced per project, team and offering.



API-first, GUI included

Everything via REST; HomePort web UI, SDKs and config-management modules.

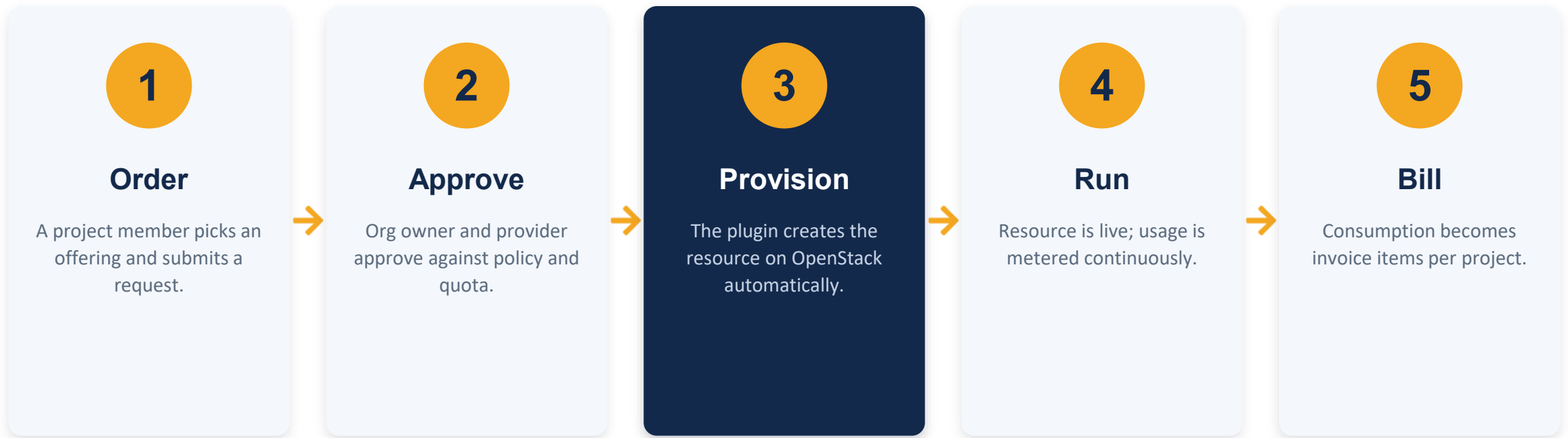
Architecture: a control plane, not a data plane



Waldur drives backends through agents and APIs. Your data and your AI jobs never pass through it.

From request to running OpenStack resource

Self-service replaces the ticket queue — with approvals and metering built in.



Turning GPUs into governed allocations

AI capacity is scarce and expensive. Waldur makes it shareable on your terms.



GPU offerings

Publish GPU flavours and HPC partitions as orderable marketplace items.



Quota-based limits

Per-project and per-user limits on larger allocations, enforced at the backend.



Service accounts

Scoped, time-limited shared credentials for pipelines and agents.



Storage allocations

Request and visualise project storage in its own terms.

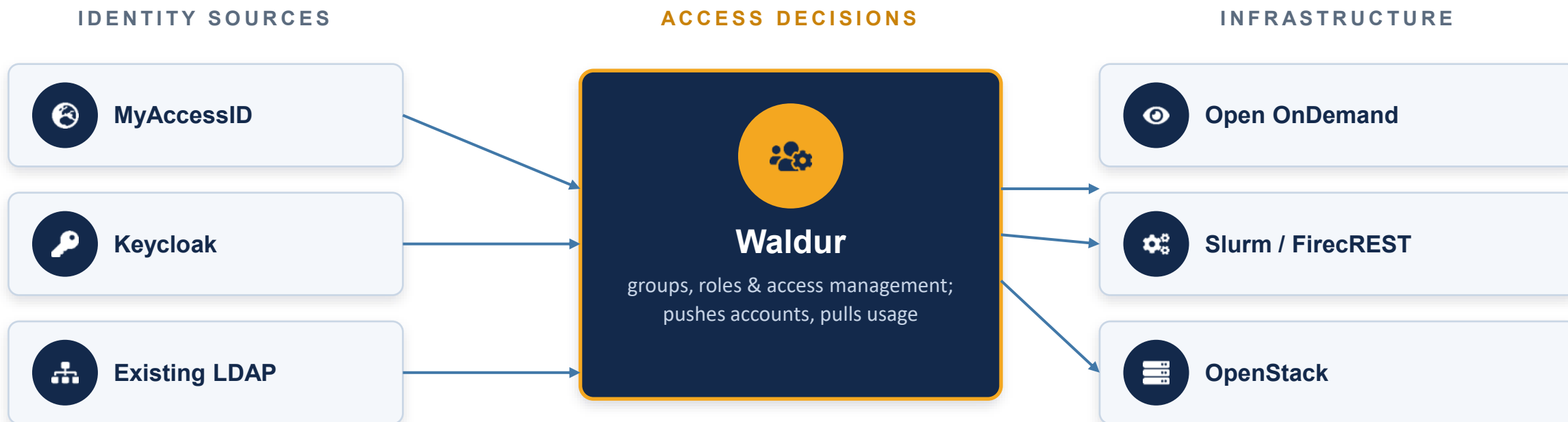


Also emerging — inference-as-a-service: order a model, get an OpenAI-compatible endpoint and a browser playground (provider-agnostic).

WHO GETS IN

Identity & access: zero-trust, federated

Waldur becomes the place where group membership and access are decided — not each backend.



Federated login in, consistent access out — across HPC, web portals and OpenStack alike.

Governance you can actually prove

The controls that turn a shared cluster into an accountable, compliant service.



Policy engine

Automated limits and actions on projects and resources.



Compliance checklists

Configurable questions to clear projects, users and proposals.



Terms of service

Offering-specific ToS with explicit, revocable consent.



Call & access review

Manual or automated review processes for requesting access.

IN PRODUCTION

Already running real sovereign infrastructure

Riigipilv

Estonian Government Cloud

National public-sector cloud on OpenStack, self-service via Waldur.

EuroHPC

Federation platform

Shared access portal across European supercomputers.

CSCS

Switzerland

National supercomputing centre.

Isambard-AI

Bristol, UK

One of the UK's largest AI supercomputers.

Cambridge

Dawn

Research GPU computing at scale.

Sovereign IaaS, plus a platform on top

OpenStack keeps your AI compute sovereign. Waldur makes it self-service, governed and accountable — open source, no lock-in.



MIT licensed



Policy inside



Orchestration outside

Ilja Livenson
OpenNode OÜ

opennodecloud.com
waldur.com · MIT on GitHub

Stay for the panel → “Two sides of the same cloud”

FURTHER READING

AI on Waldur & sovereign research infrastructure

1

Beyond Pre-Training: The Full Lifecycle of Foundation Models on HPC Systems

Conciatore, Oggian, Da Forno, Schuppli, Tissieres, VandeVondele & Martinasso (CSCS) [Cray User Group 2026](#) · [arXiv:2604.12599](#)

Frames national HPC as the backbone of sovereign AI and the EuroHPC AI Factories — pre-training, fine-tuning and inference.

2

Isambard-AI: a leadership class supercomputer optimised specifically for Artificial Intelligence

McIntosh-Smith, Alam & Woods (University of Bristol) [Cray User Group 2025](#) · [doi:10.1145/3725789.3725794](#)

Describes the MyAccessID → Waldur single sign-on and access workflow for the Isambard-AI portal.

3

Federated Single Sign-On and Zero Trust Co-design for AI and HPC Digital Research Infrastructures

Alam et al. [SC24-W Workshops](#) · [doi:10.1109/SCW63240.2024.00220](#)

The zero-trust, federated-identity design that Waldur implements across HPC and cloud.

What Waldur drives in OpenStack

One OpenStack.Tenant offering, and users self-serve the full IaaS surface through Keystone, Nova, Cinder, Neutron and Glance.



Keystone

Identity

- Projects / tenants
- Users & credentials
- Quota sync



Nova

Compute

- Instances & flavours
- GPU flavours
- Console, affinity groups



Cinder

Block storage

- Volumes & types
- Snapshots
- Backups



Neutron

Networking

- Networks, routers
- Security groups, FIPs
- Load balancers (Octavia)



Glance

Images

- OS images
- Custom image upload
- Per-tenant limits



Quotas map 1:1 — cores, RAM, storage, instances, volumes, floating IPs — enforced in both directions.

Provision via web UI, REST, Python SDK, Ansible (waldur.openstack.org) or Terraform.

OpenStack integration, in detail



Load balancing (Octavia)

- Octavia via openstacksdk — OVN provider only
- Listeners & pools: TCP / UDP; algorithm SOURCE_IP_PORT
- LBs, listeners, pools, members, health monitors
- Floating IP + security groups on the VIP port
- Background reconcile every 2 hours



Authentication & sessions

- Password or v3 application credential
- App credential: username = ID, password = secret
- Scoped & revocable — recommended for production
- Keystone sessions cached ~10 h, auto-refreshed



Storage modes & quotas

- Fixed mode: one shared storage quota
- Dynamic mode: per volume-type quotas (SSD / HDD)
- Quotas reconciled in both directions
- Defaults e.g. 20 cores, 50 GB RAM, 30 instances, 1 TB



Networks & sync

- External networks: provider-global or RBAC-exposed
- Router gateway: two-tier perms (SNAT / fixed IPs)
- Periodic pulls keep Waldur ↔ OpenStack consistent
- Clients: Keystone v3 · Nova v2.19 · Cinder v3 · Glance v2 · Neutron v2.0

XAAS FROM IDEAS
TO IMPACT.

**VIELEN DANK
FÜR IHRE ZEIT!**

XAAS GmbH & Co. KG
Kobestraße 7, 20457 Hamburg
T: (+49) 40 / 22 86 90 40
E: hallo@xaas-it.com

