

e-INFRA Kubernetes

Infra ve stínu AI

Managed platforma K8s

- Centrálně spravovaná platforma pro spouštění kontejnerů
 -  Uživatel řeší **vlastní aplikace**, ne infrastrukturu
 -  Integrace úložišť, DNS, certifikátů
 -  Předpřipravené aplikace (operátory)
 - PostgreSQL (**CloudnativePG**), **OpenSearch**, **Spark**, Redis (**Dragonfly**)
-  Nelze spouštět kontejnery jako root
-  Uživatel nemá admin práva K8s clusteru (nelze instalovat CRD, ClusterRole)

Co je nového?

- Několik nových uzlů včetně GPU NVIDIA RTX Pro 6000
- Aktualizace služeb – JupyterHub, Foldify
- Velký rozvoj **AI**
- Chystáme se na změnu některých komponent
- Nasazení **CephFS**
- Stabilní provoz

Infrastruktura dnes

- **Kuba cluster**
 - 51 fyzických uzlů – baremetal bez virtualizace
 - 3000 CPU jader
 - **58 GPU** – A40, A100, H100, L4, RTX Pro 6000
 - NVME storage, některé uzly až 40TB
 - Sít':
 - 2x 10Gb, 100Gb
 - InfiniBand – 100, 200, 400Gb
- **Kubas cluster** (sensitive)
 - 19 fyzických uzlů podobné konfigurace, bez InfiniBandu

K8s clustery

- **Kuba** cluster (<https://kuba-lb.cloud.e-infra.cz>)
 - Veřejný cluster pro kohokoliv s Metacentrum účtem
 - Pouze jedna lokalita v Brně
- **Kubh** cluster (<https://kubh-lb.cerit-sc.cz>)
 - Uzavřený cluster
 - 3 lokality v Brně
- **NRP** cluster (<https://app1.nrp1.du.cesnet.cz>)
 - Cluster národní repozitářové platformy – pro repozitáře
 - 2 lokality (Ostrava, Ústí nad Labem) – přibydou další (možná, podle trhu)
- **Kubas** cluster (<https://kubas-lb.cloud.trusted.e-infra.cz>)
 - Cluster pro zpracování citlivých dat, speciální onboarding, přístup přes VPN
 - 2 lokality v Brně

Možnosti přístupu

- **Webová rozhraní**

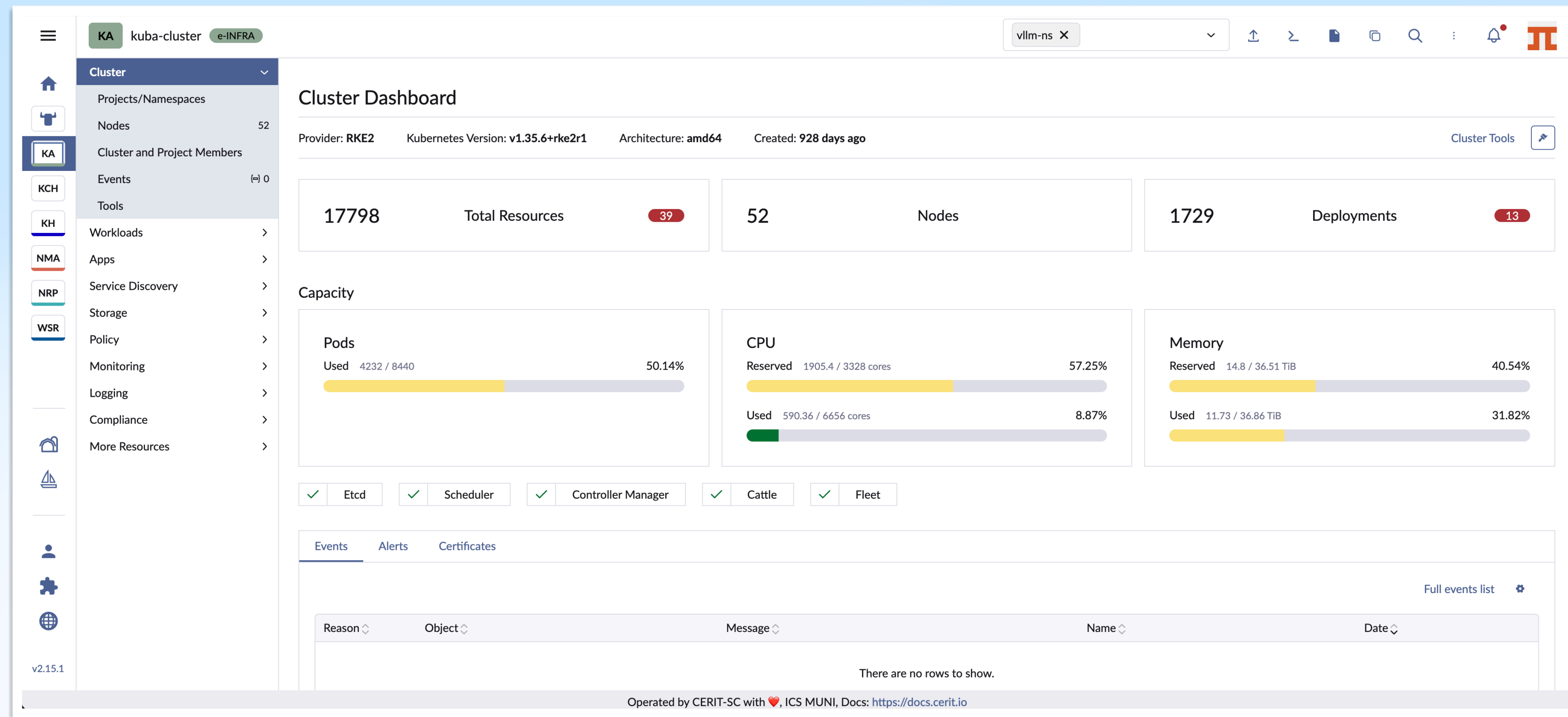
- <https://rancher.cloud.e-infra.cz>
- JupyterHub
- Foldify
- AI

- **Kubernetes API**

- kubectl, helm

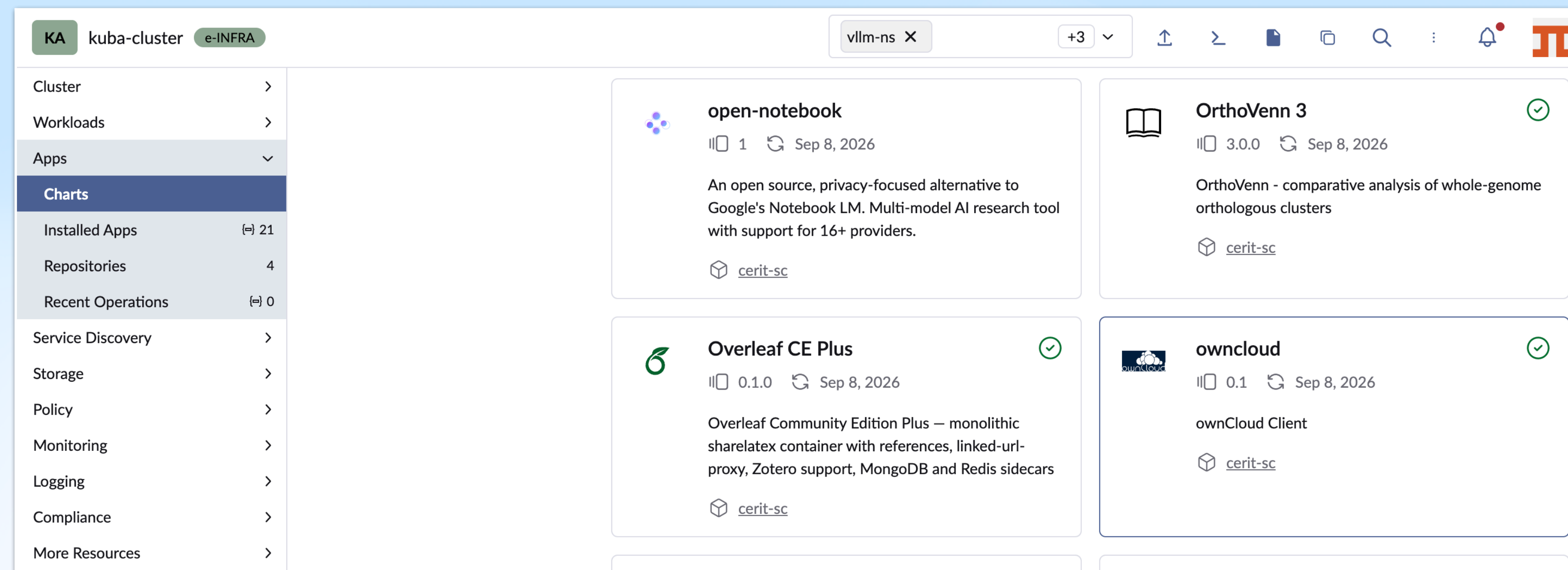
- **Automatizace**

- GitLab agent
- ArgoCD/Workflows
- Docker Runner (v přípravě)



Nové a rozšířené služby Rancher

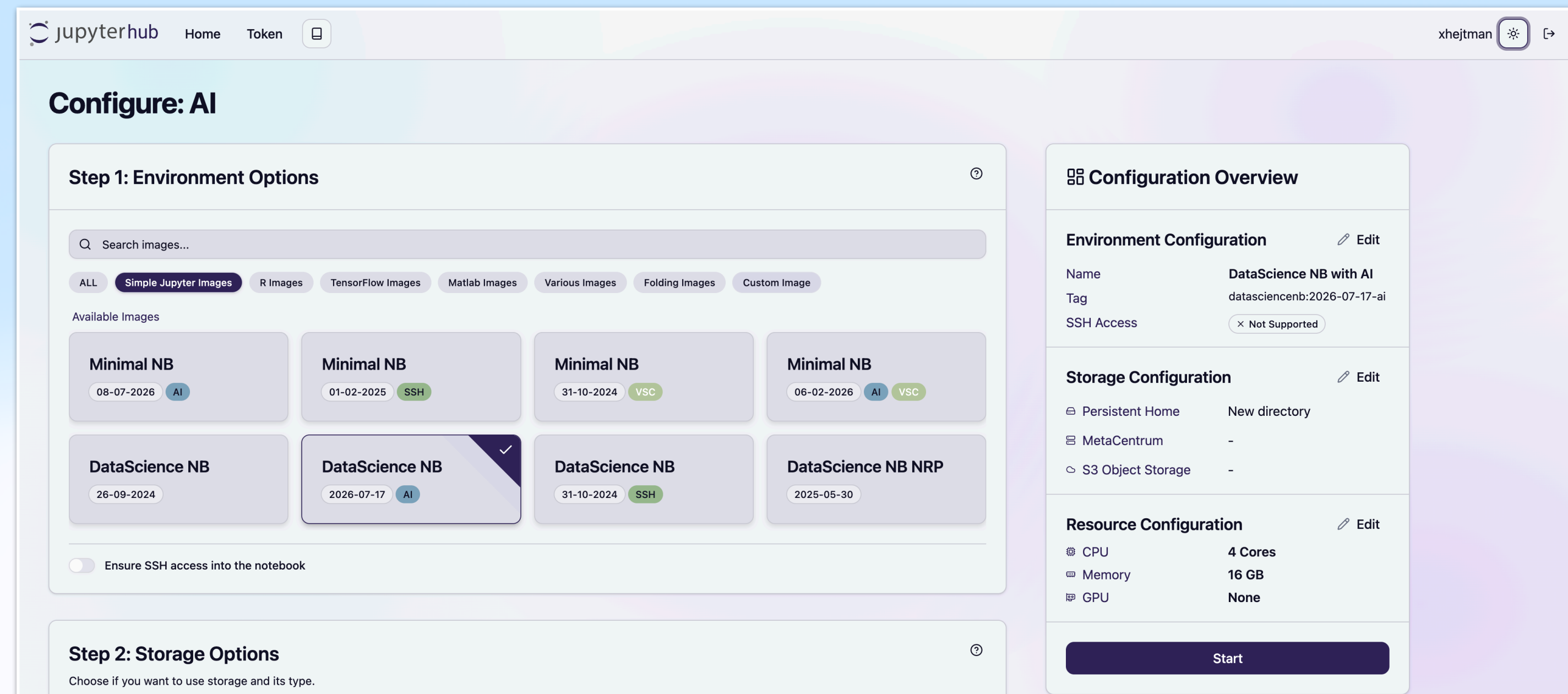
- Nově přidané aplikace
 - Open Notebook
 - OrthoVenn 3
 - Overleaf



Rozšířené služby

JupyterHub – <https://hub.cloud.e-infra.cz>

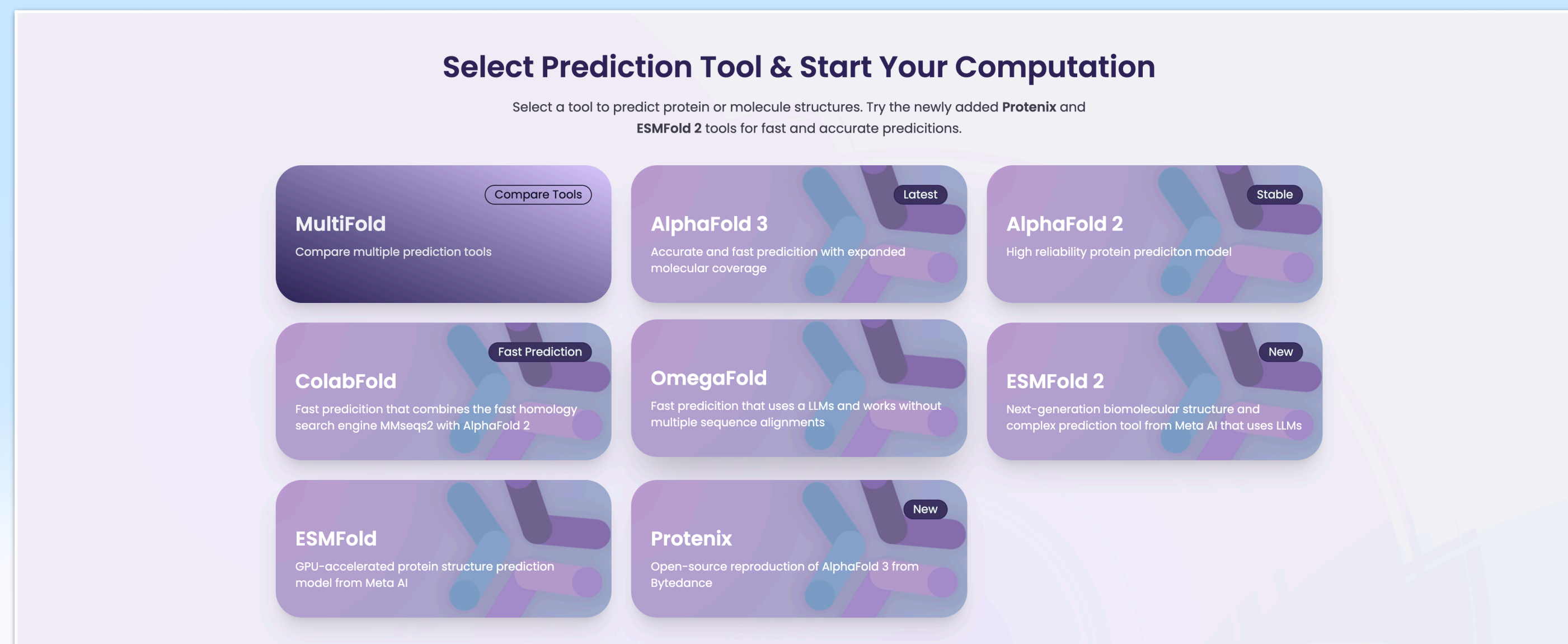
- Nová verze rozhraní
- Integrace AI
 - DataScience NB AI
 - Claude code
 - OpenCode
 - **RStudio** 4.6.1 s nativní AI – Posit Assistant konečně funguje odhlášení
- Integrace s **NRP**
 - NRP klient v DataScience NB
- Podpora dalších obrazů
 - RStudio, Google Colab, Matlab, Open Notebook



Rozšířené služby

Foldify – <https://foldify.cloud.e-infra.cz>

- Nově přidáno
 - **ESMFold2**
 - **Protenix**
 - **MultiFold**
 - Porovnání výsledků různých nástrojů
- Implementovaný frontový systém
- Využívá mmseq search (interně <https://mmseqs.cerit-sc.cz>)



AI služby

- **Chat**

- <https://chat.ai.e-infra.cz>

- **LLM API**

- <https://llm.ai.e-infra.cz>

- **Monitoring**

- Status: <https://llm.ai.e-infra.cz/status>
- Usage: <https://llm.ai.e-infra.cz/usage>

- Podrobněji v další prezentaci

API a automatizace

- Rancher jako API proxy
- Podpora pro kubectl, helm
- Na žádost ArgoCD/ArgoWorkflows – optimalizace oprávnění
- Gitlab Agent CI/CD
- V přípravě Docker runner

Změny v komponentách

Ingress

- Přejechod z **NGINX** Ingress na **Traefik** Ingress
 - Semi-kompatibilní NGINX Ingress varianta (ingressClassName: **nginx-traefik**)
 - Gateway API (**Gateway** + **HTTPRoute** manifesty)
 - IngressRoute
- **Změna veřejného jména**
 - <https://kuba-lb.cloud.e-infra.cz>
(místo dřívějšího <https://kuba-pub.cerit-sc.cz>)

Změny v komponentách

Postgres/CloudnativePG

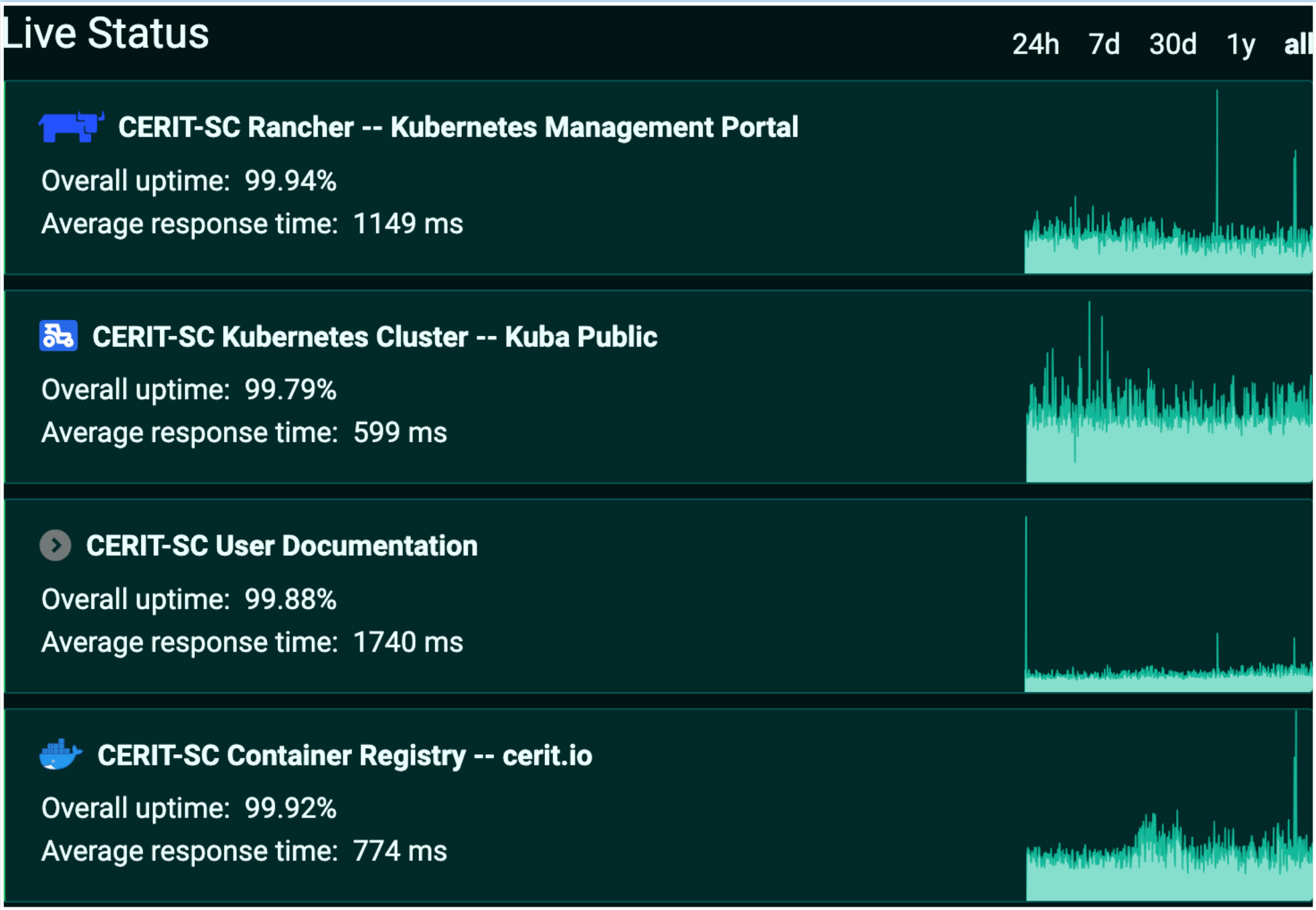
- Minimální podporovaná verze Postgres 14, do konce roku pravděpodobně min. 15
- Aktualizace major verze podporovaná – skok minimálně na verzi 16.
- Změna systému zálohování databáze
 - Přejechod z Barman Cloud na Barman Cloud Plugin
 - Funkčně stejné, mechanismus jiný
 - CloudnativePG stále odkládá úplné ukončení podpory staré verze
 - *Updated the deprecation notice for native (in-tree) Barman Cloud support to reflect that it will now be removed in CloudNativePG 1.31.0, rather than 1.30.0. **Users are still encouraged to migrate to the Barman Cloud Plugin.***

Změny v komponentách

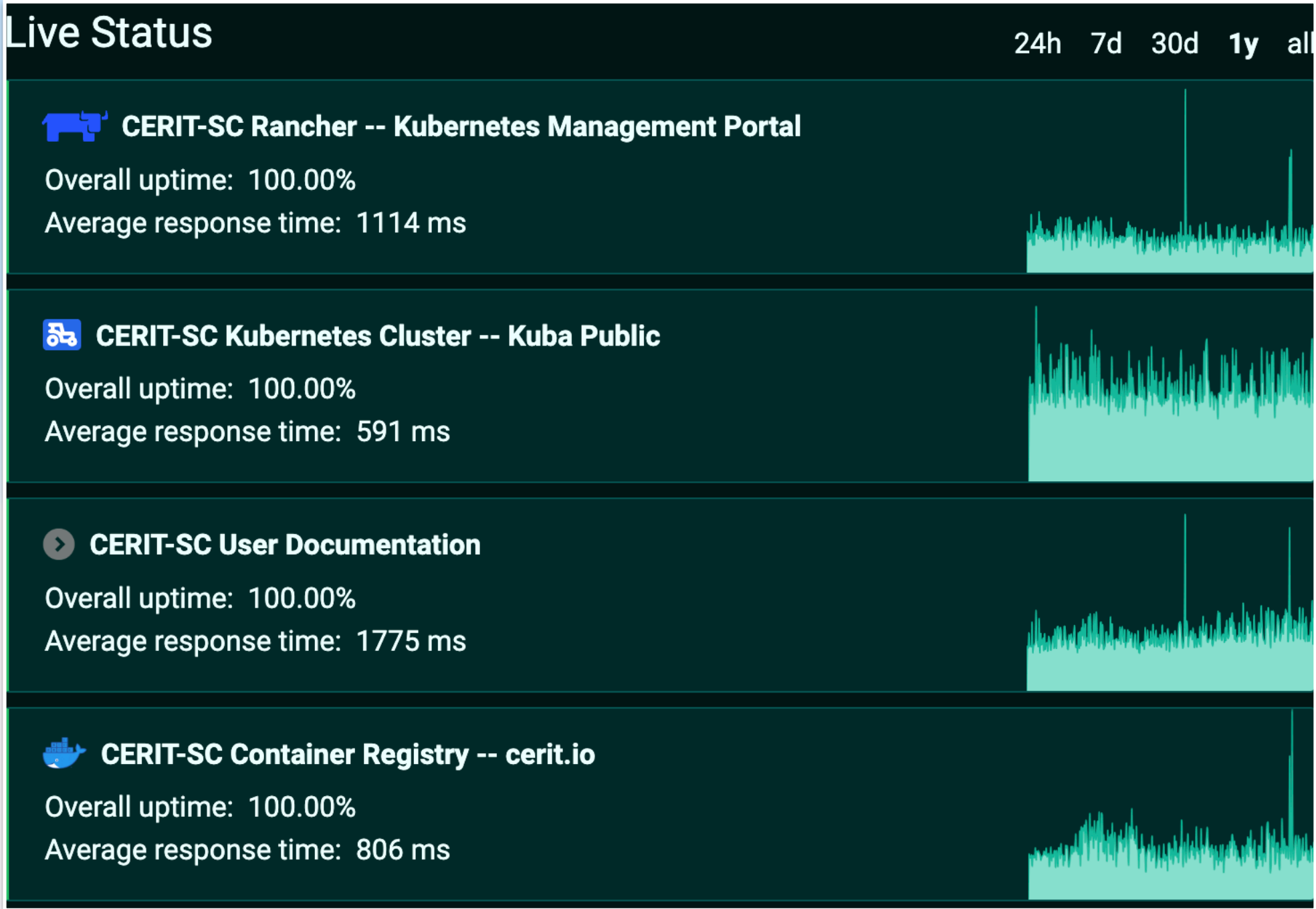
Úložiště

- Aktuálně v provozu
 - NFS (**nfs-csi**), all flash úložiště, částečně dosluhující
 - Beegfs (**beegfs-csi**), all flash úložiště, experiment, ve kterém se nepokračuje
 - Lokální disky (**zfs-csi**), all flash, vázané na konkrétní uzel, rychlé
- Nově
 - CephFS/S3 (**rook-cephfs**), all flash, dostupné na kub-b/c/d uzlech
 - Podpora snapshotů
 - Respektuje požadovanou velikost (narozdíl od nfs-csi)

Stabilita infrastruktury



Za dobu provozu



Za poslední rok

Velký bratr se dívá

Monitoring využití zdrojů

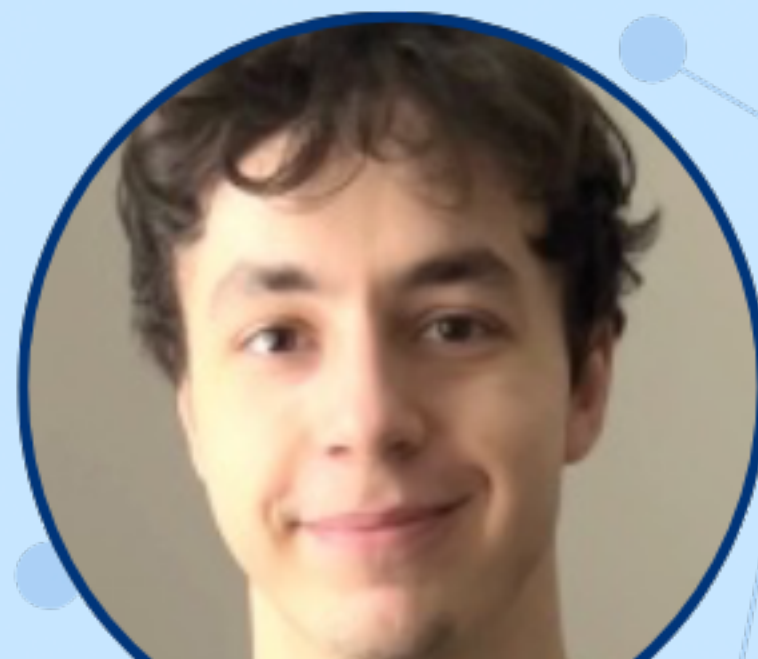
- Sledujeme, co se na infrastruktuře skutečně děje, jak je využita
- **AI Karel**
 - Sada skriptů + LLM
 - Sleduje
 - Hledá
 - Analyzuje proč
 - Hlásí

GPU waste						
A GPU held at zero utilization is not a resource in reserve; it is a promise made by one workload against every other workload that might have been. Twelve pods below report n/a — they keep the card and give nothing back.						
• Pods evaluated (age >= 12h): 32 • Excluded (age < 12h): 8 • Top 30 GPU wasters listed below, ranked by increasing real GPU usage (worst waste first). • Pods marked NO GPU USAGE are holding a GPU while reporting n/a (zero utilization).						
#	Namespace	Pod	Real GPU Usage	Age	GPU Name	GPU Used
1	bajger-ns	claude-claude-cli-5f88977f97-n927v	n/a	20h39m	NVIDIA A40	NO GPU USAGE
2	bohac-ns	vllm-qwen32-58bccd97f6-4zznm	n/a	20h28m	NVIDIA A40	NO GPU USAGE
3	filipovic-ns	vm-example-3-0	n/a	2d18h56m	NVIDIA H100 NVL	NO GPU USAGE
4	jupyterhub-mariapinillav-prod-ns	jupyter-mariapinillav-r42-df52e231	n/a	1d1h50m	NVIDIA A100 80GB PCIe-MIG	NO GPU USAGE
5	jupyterhub-martav-prod-ns	jupyter-martav-cloudecod-0030b59a	n/a	21h44m	NVIDIA A100 80GB PCIe-MIG	NO GPU USAGE
6	jupyterhub-ondryx-prod-ns	jupyter-ondryx-ai-training-a61a8e1c	n/a	5d1h21m	NVIDIA H100 NVL	NO GPU USAGE
7	mdfind-ns	prod-md-find-embedder2-58f8cf74f4-8nnxr	n/a	20h28m	NVIDIA RTX PRO 6000 Blackwell Server Edition	NO GPU USAGE

Co dál?

- Další integrace AI
 - Do správy infrastruktury
 - Směrem k uživatelům
- Aktualizace komponent
 - GPU: **DRA** mechanismus alokace
- Nový hardware patrně nejdříve v 2028
- Cíl: aby uživatel řešil svůj výzkum, ne provoz infrastruktury.

Tým



SEBASTIÁN BENCSÍK



ROMANA ĎURÁČIOVÁ



KRISTIÁN KOVÁČ



JANA TULIKOVÁ



JOSEF HANDL



LUKÁŠ MARTIŠKA



ROBIN DURNÍK



KRYŠTOF ALBRECHT



ZDENĚK SALVET



MICHAL JÁL

Děkuji za pozornost

k8s@cerit-sc.cz

<https://docs.cerit.io>

<https://blog.e-infra.cz>