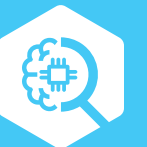


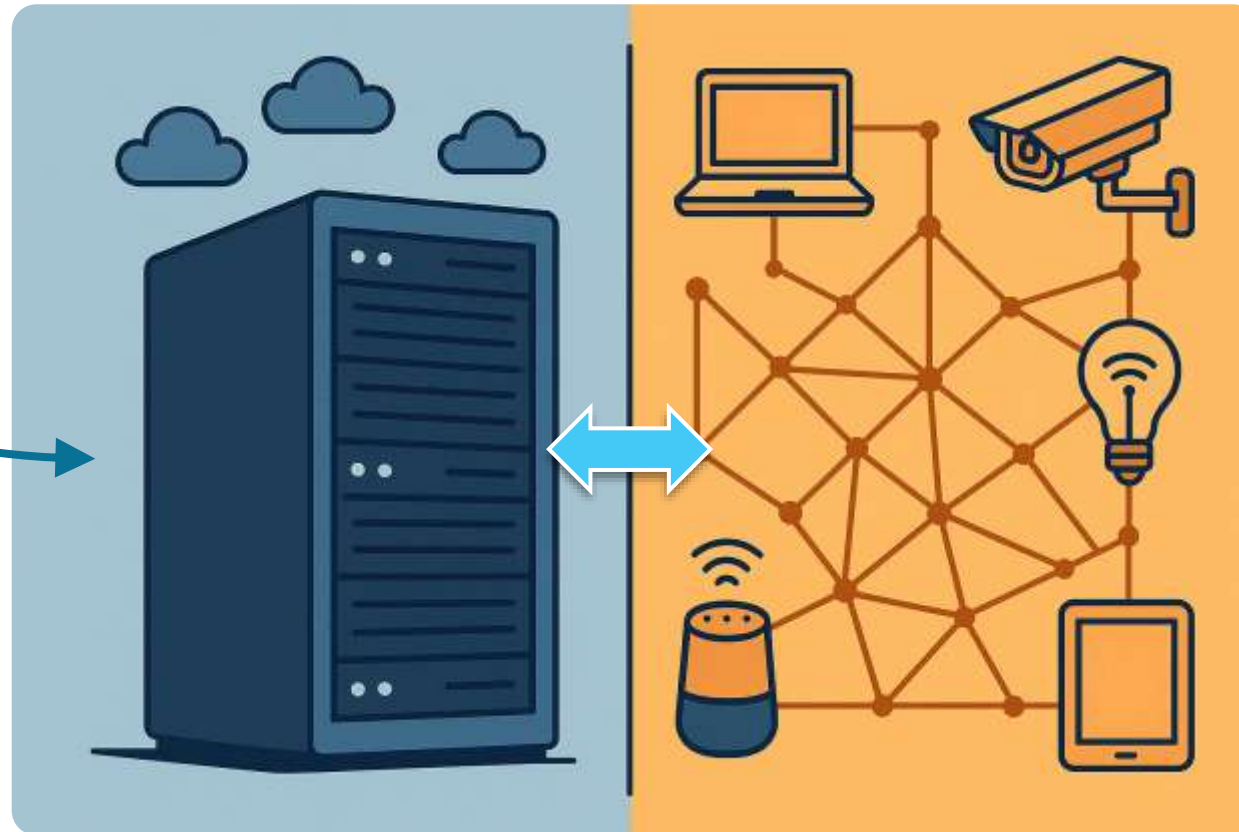
howest

Kubeflow

MLOps4ECM



Introductie



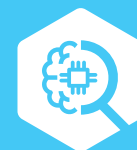
Gecentraliseerd

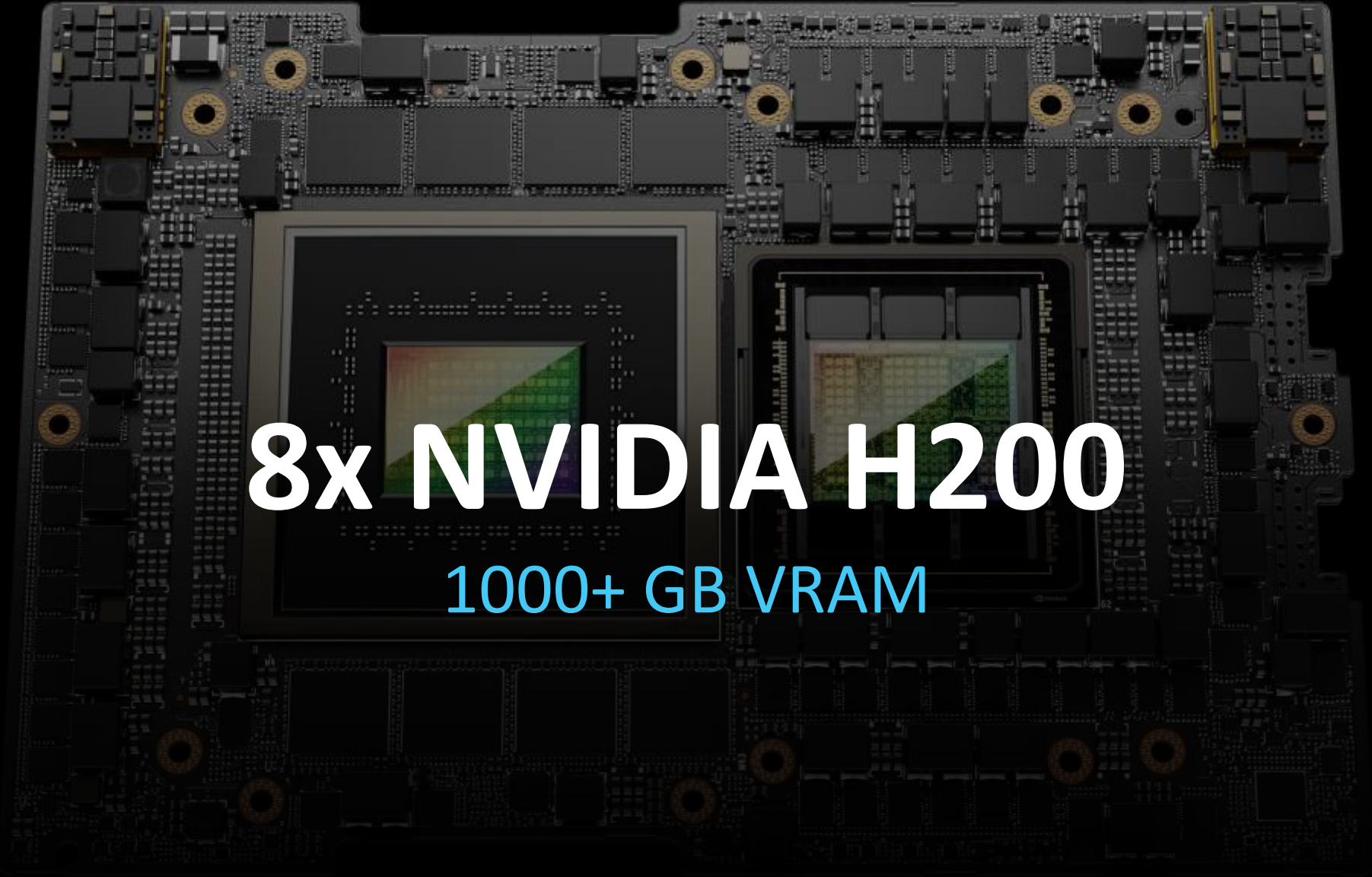
Gedecentraliseerd

howest

Onze hardware

Howest AI Lab infrastructuur



A top-down view of a server-grade GPU, specifically an NVIDIA H200, mounted on a dark-colored printed circuit board (PCB). The GPU is a large, square component with a central die and numerous peripheral components like capacitors and smaller chips. The PCB has several circular mounting holes and various connectors along its edges. The background is dark, making the components stand out.

8x NVIDIA H200

1000+ GB VRAM

Infrastructuur



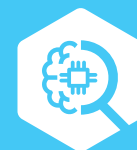
GPU	8x NVIDIA H200 SMX (141GB)
CPU	2x Intel Xeon Platinum 8562Y+ (128 threads)
RAM	2TB (DDR5)
Storage	25TB in totaal



howest

Onze software

Howest AI Lab infrastructuur



Cloud native landscape

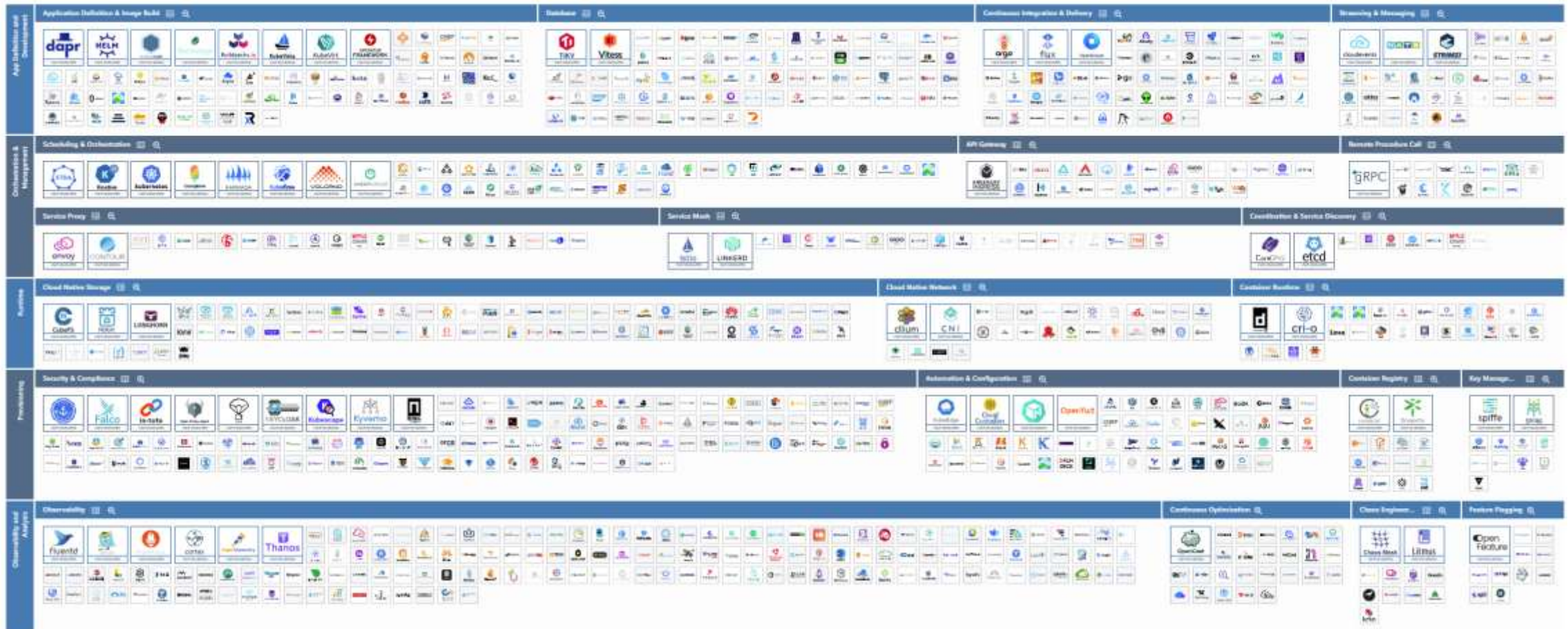


**CLOUD NATIVE
LANDSCAPE**

landscape.cncf.io



Cloud native landscape



Bevat zeker nuttige zaken voor IoT, edge-deployments, MLOps, ...

CNCF onderdelen



Application Definition & Development

Orchestration & Management

Runtimes

Provisioning

Observability & Analysis



Scheduling & Orchestration



Orchestration & Management

Scheduling & Orchestration

 CNCf GRADUATED	 CNCf GRADUATED	 CNCf GRADUATED	 CNCf INCUBATING	 CNCf INCUBATING	 CNCf INCUBATING	 Amazon ECS						
 CNCf INCUBATING	 CNCf INCUBATING	 ARMADA	 Azure Service Fabric	 capsule	 移动云 CNP	 中移磐基	 Clusternet	 Clusterpedia	 SWARM	 DolphinScheduler		
 kube-green	 kube-rs	 KubeAdmiral	 KubeWharf	 ERASER	 FLUID	 HAMi	 iSSCloud	 KCP	 kOS	 kORDENT	 kestra	 koordinator
 PREFECT	 SERVERLESS DEVS	 StackStorm	 upbound	 Katalyst	 KUBESTELLAR	 Kured	 slurm	 Nomad	 Open Cluster Management	 OPEN FUNCTION	 Open Nebula	



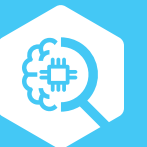
CNCF niveau's

1. **Sandbox** = ontwikkelingsfase, experimenteel, zoeken funding, ...
2. **Incubating** = worden onderhouden, ontstaan van community, ...
3. **Graduated** = betrouwbaar, vaste waarde, sterke reputatie, ...



howest

Kubernetes



Kubernetes



Kubernetes is een workload orchestrator:

- Beheert resources (CPU, geheugen, ...)
- Automatisch beheren, schalen en verdelen van workloads over 1 of meerdere computers



Waarom Kubernetes?



- Open-source
- **Integratie met NVIDIA** (hardware & software)
- Ondersteunt **GPU sharing**
- Schaalbaarheid
- **Uitbreidbaar**
- ...

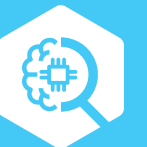


Kubernetes is overall



howest

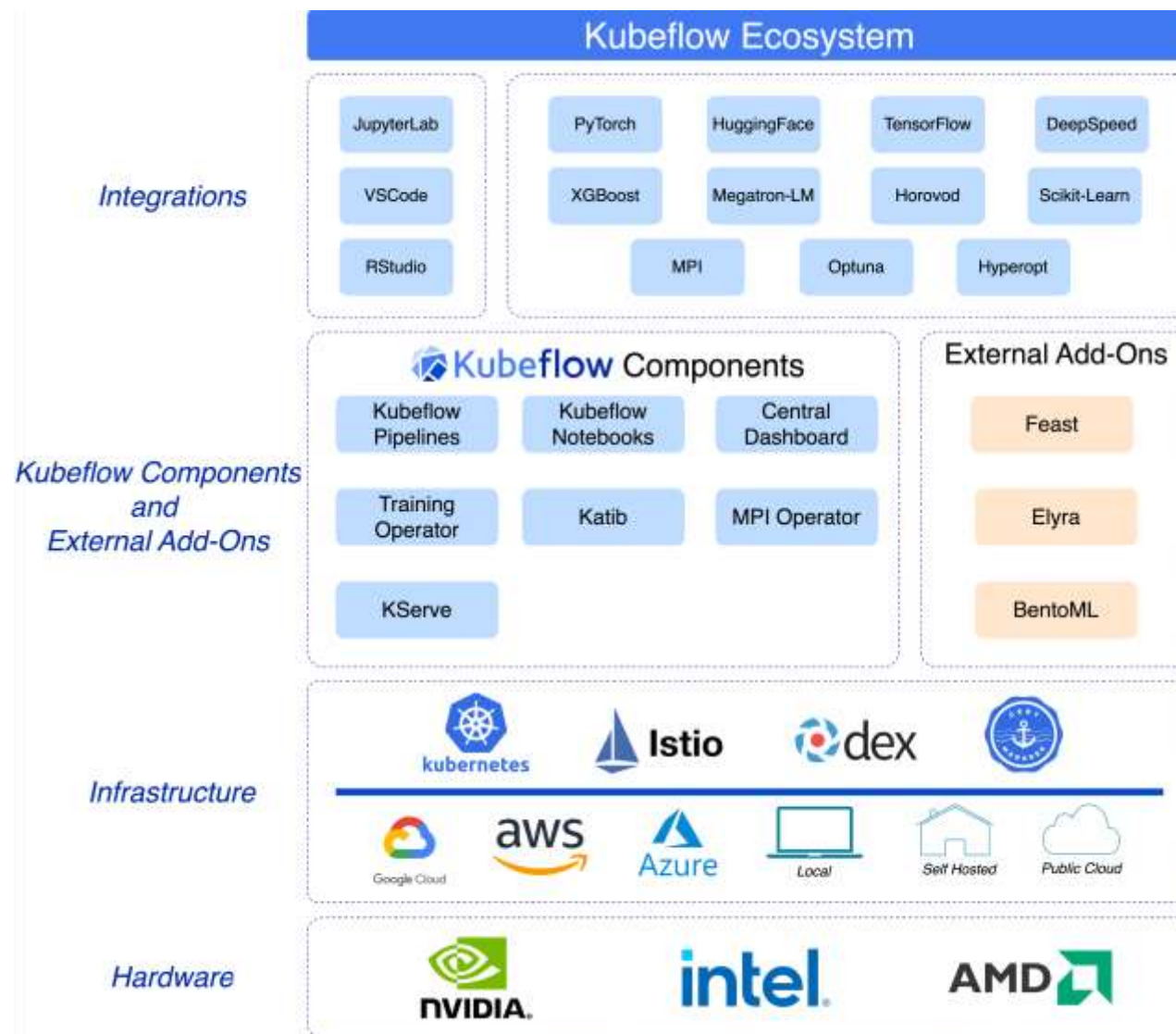
Kubeflow



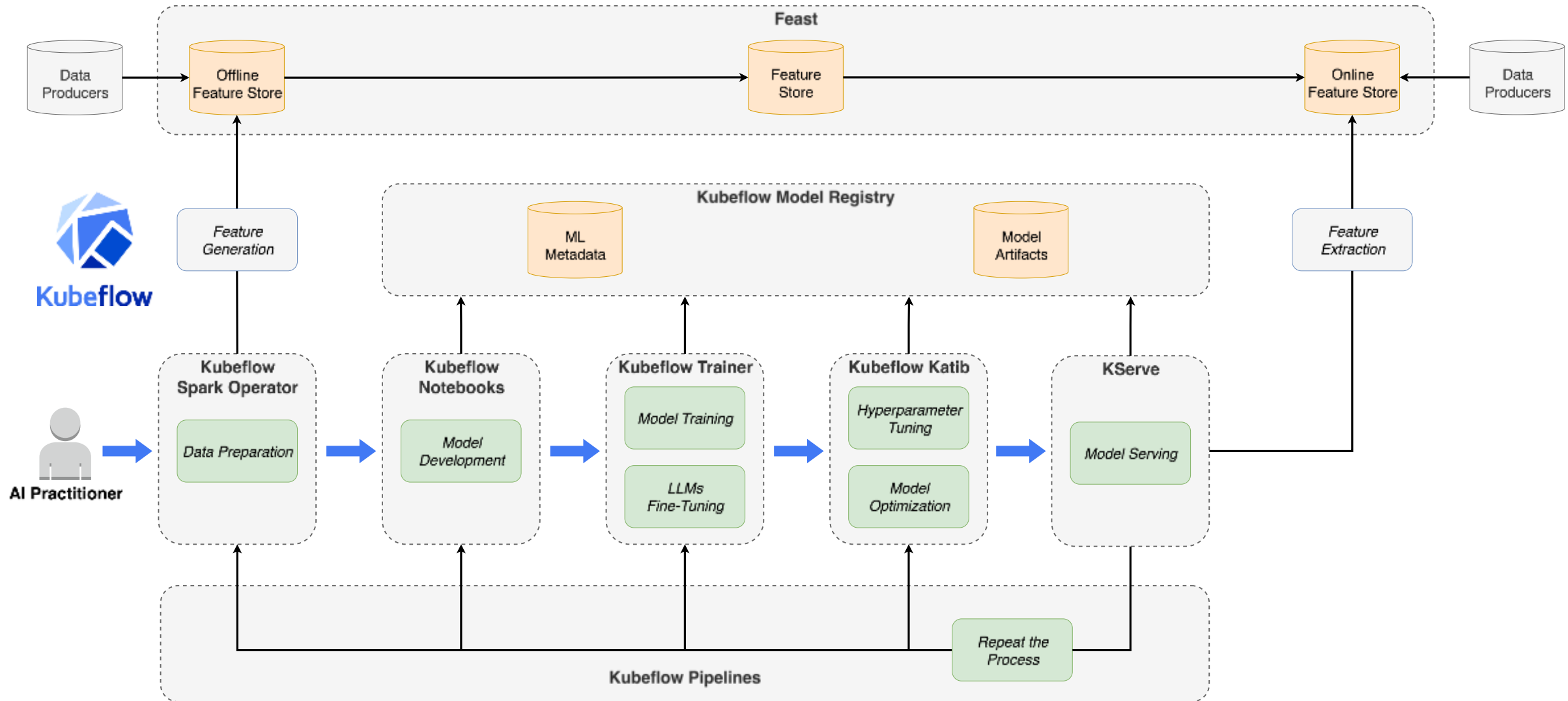
Kubeflow



Open-source machine learning platform dat draait op **Kubernetes**.

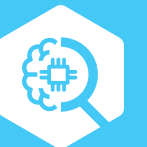


Kubeflow working



howest

Kubeflow Pipelines





Kubeflow Pipelines

1. Gebruik de KFP library
2. Exporteer naar YAML

```
complete-pipeline.py x ! kfp_214_minimal.yaml
complete-pipeline.py > ...
1  from typing import NamedTuple
2  from kfp import dsl, compiler
3  from kfp.dsl import component, Input, Output, Dataset, Metrics, Model
4
5  # ----- components -----
6
7  @component(base_image="python:3.11-slim")
8  def make_numbers(n: int, out_list: Output[Dataset], label: str = "run"):
9      with open(out_list.path, "w") as f:
10         for i in range(n):
11             f.write(f"{i}\n")
12         print(f"[make_numbers] label={label}, n={n}")
13
14  @component(base_image="python:3.11-slim")
15  def filter_evens(in_list: Input[Dataset], out_list: Output[Dataset], m: Output[Metrics]):
16      kept = []
17      with open(in_list.path) as f:
18         for line in f:
19             x = int(line.strip())
20             if x % 2 == 0:
21                 kept.append(x)
22      with open(out_list.path, "w") as f:
23         for x in kept:
24             f.write(f"{x}\n")
25      m.log_metric("even_count", len(kept))
26
```

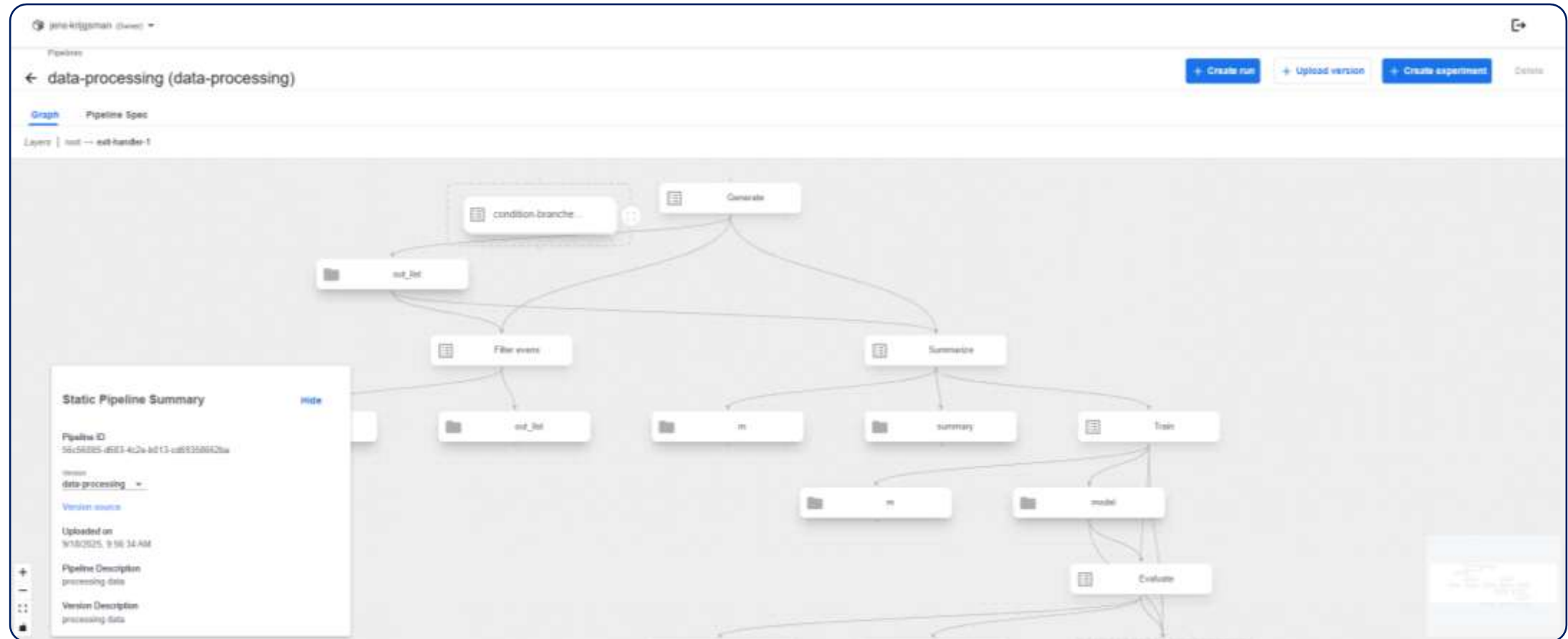
Kubeflow Dashboard



1. Gebruik de KFP library
2. Exporteer naar YAML
3. Upload op Kubeflow

A screenshot of the Kubeflow Dashboard's 'New Pipeline' form. The interface has a dark blue sidebar on the left with navigation links: Home, Notebooks, TensorBoards, Volumes, Katib Experiments, KServe Endpoints, Pipelines (highlighted), Experiments, Runs, Recurring Runs, Artifacts, Executions, Profiles, and Manage Contributors. The main content area is white and titled 'New Pipeline' with a back arrow. It shows the user 'jens-krijgsman (Owner)'. Under 'Pipeline Versions', there are two radio buttons: 'Create a new pipeline' (selected) and 'Create a new pipeline version under an existing pipeline'. Below this, there are two radio buttons for 'Private' (selected) and 'Shared'. The 'Upload pipeline with the specified package.' section contains a 'Pipeline Name' field with the value 'data-processing' and a 'Pipeline Description' field with the value 'sample-data-pipeline'. A note says 'Choose a pipeline package file from your computer, and give the pipeline a unique name. You can also drag and drop the file here.' and a link to 'Compile Pipeline Documentation'. There are two options for uploading: 'Upload a file' (selected) with a file input field containing 'kfp_214_minimal.yaml' and a 'Choose file' button, and 'Import by url' with a 'Package Url' input field. At the bottom, a 'Code Source' field contains the text 'kubernetes data-processing pod'.

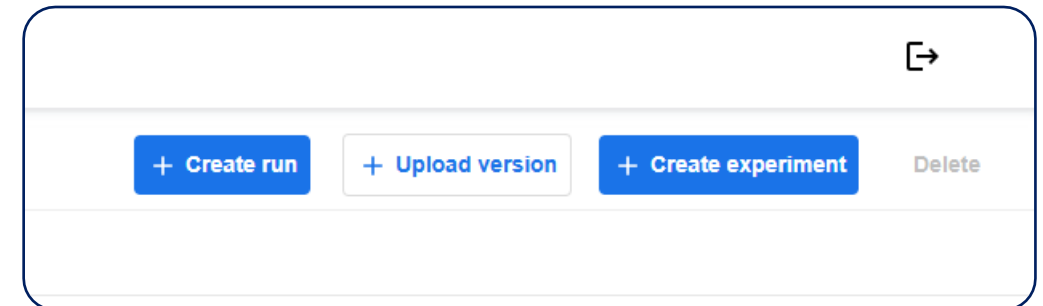
Kubeflow Pipelines





Kubeflow Pipelines

1. Bekijk de pipeline **graph**
2. Version de pipeline
3. Maak een **experiment** (= groep van **runs**)
4. Maak een **run**
 - Voer een specifieke pipeline versie uit
 - Input parameters meegeven
 - Deel van experiment



Kubeflow Notebooks



 **Kubeflow**

[Home](#)
[Notebooks](#)
[TensorBoards](#)
[Volumes](#)
[Katib Experiments](#)
[KServe Endpoints](#)
[Pipelines](#)
[Profiles](#)

[Manage Contributors](#)

[GitHub](#)
[Documentation](#)

build version - dev_local

thomas-huyghebaert (ow...)

Notebooks

Filter Enter property name or value

Status	Name ↑	Type	Created at	Last activity	Image	GPUs	CPUs	Memory		
	ai-square	1	11 days ago	-	kubeflow-python-ollama:main	1	2	4.0 Gi	CONNECT	
	workshop2	1	12 days ago	-	kubeflow-python-ollama:main	1	2	4.0 Gi	CONNECT	

Items per page: 10 1 - 2 of 2

howest

Vragen?

ailab@howest.be