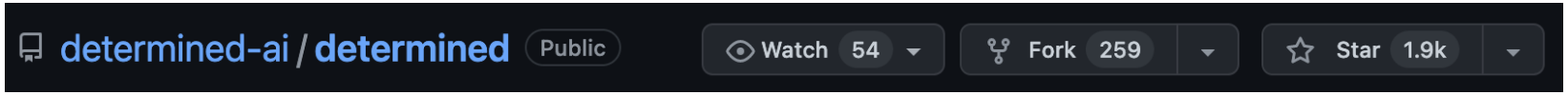
Hewlett Packard
Enterprise

HPE Machine Learning Development Environment and the Open Source Advantage

Isha Ghodgaonkar, ML Developer Advocate @ Determined AI

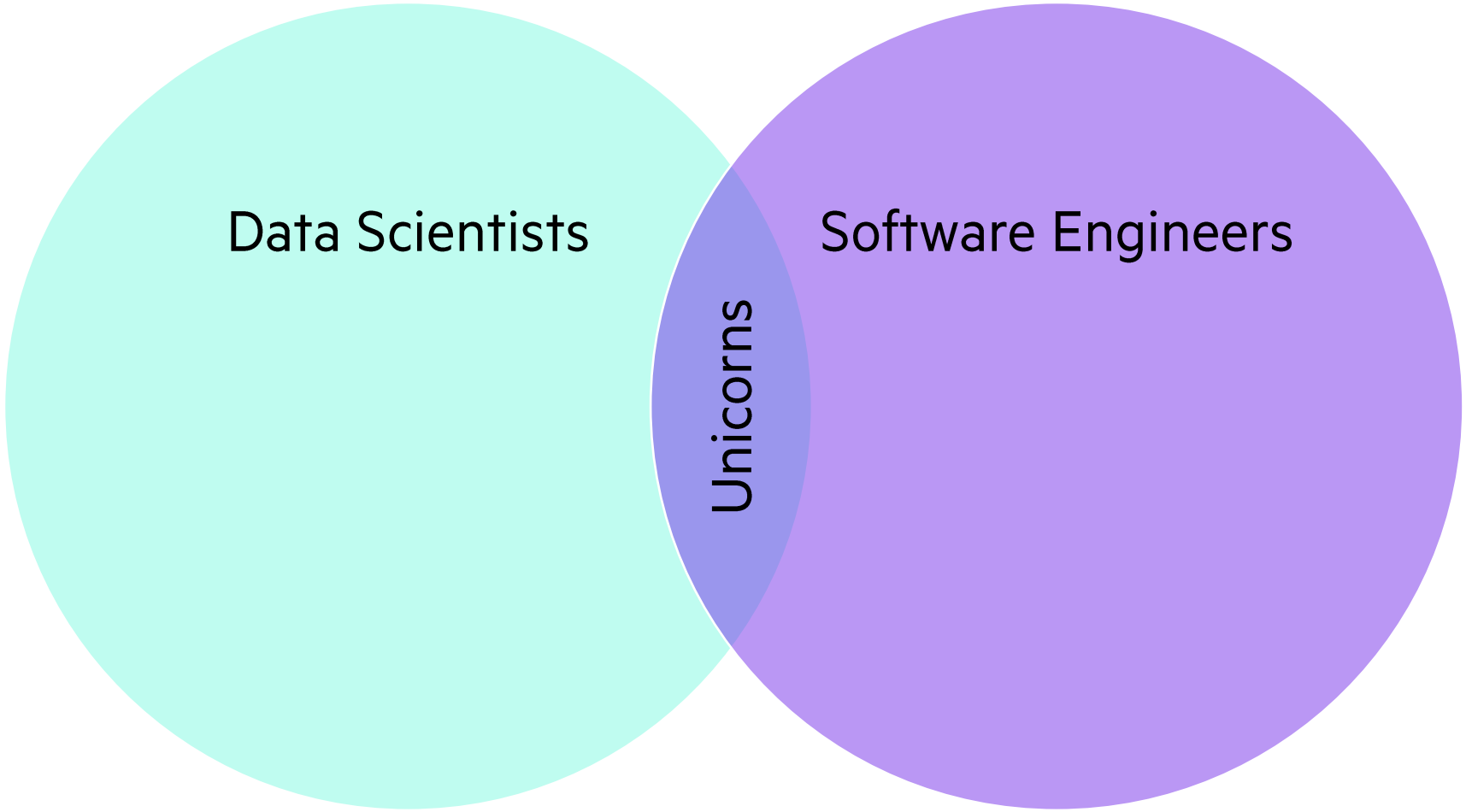
05/31/2023

Determined: Open-Source Deep Learning Training Platform



Determined Project on GitHub

Developer's guide to deploying Determined on Kubernetes



Traditional tools are built for software engineers, not data scientists

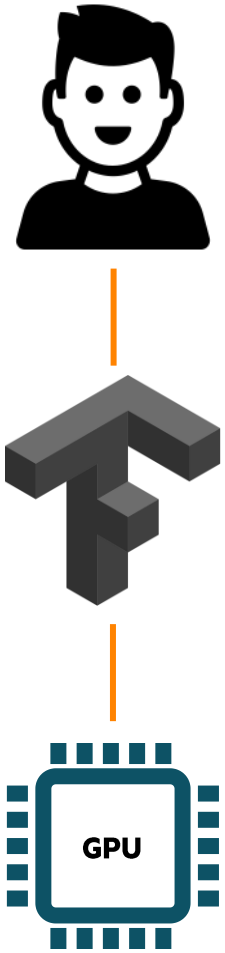
What is Machine Learning (ML) engineering?

A fundamentally different approach, compared to classical software engineering

	Software 1.0	Software 2.0
Data	None	Terabytes
Compute	Several cores	Petaflops
Hardware	CPU	GPU, TPU, etc.
Compile time	Seconds	Weeks
Accuracy	Logical (spec)	Probabilistic
Debugging	Printf	TensorBoard

Machine Learning Model Development, Early Days

On day one...



Small and simple

Relatively small experiments on single machine

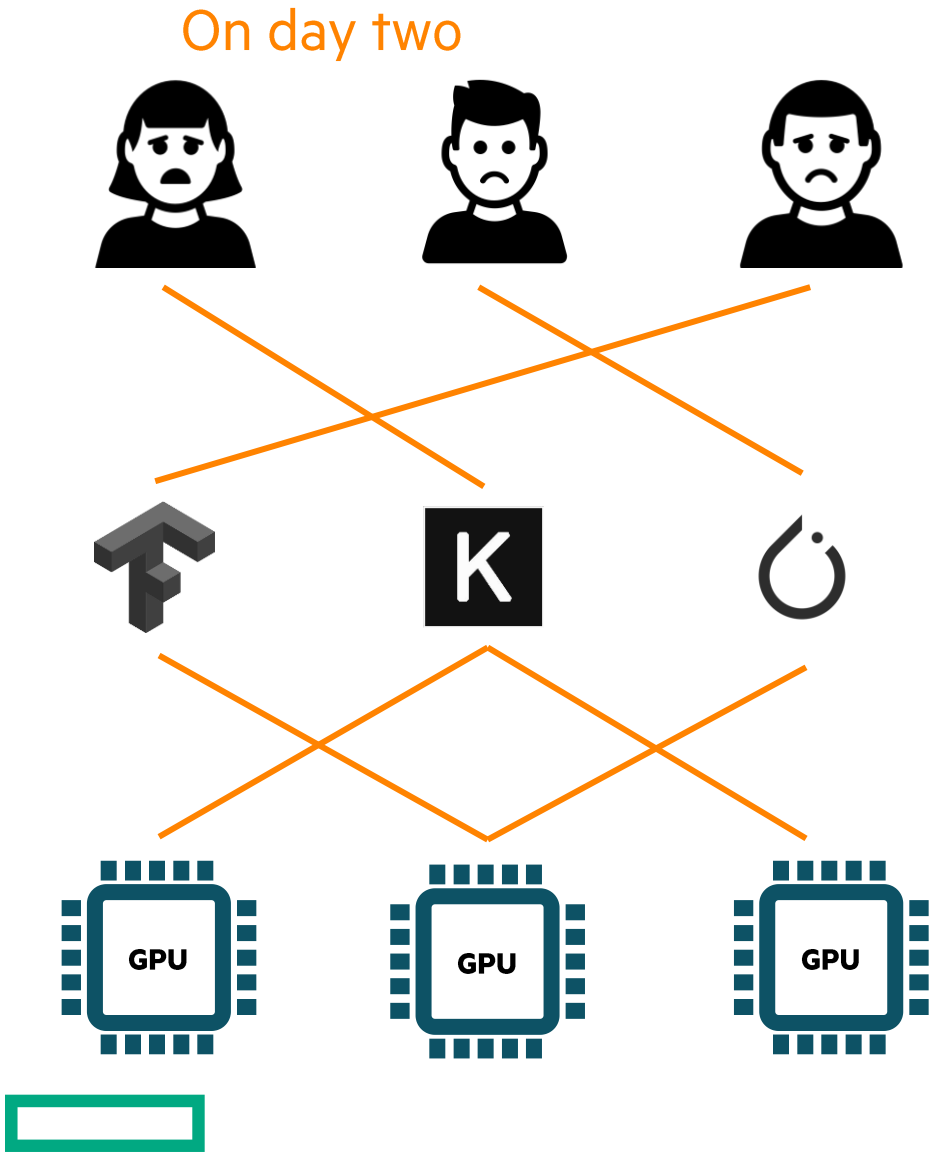
No need to share resources or productionize results

Relatively simple infrastructure to manage

Ad hoc experiment tracking is sufficient,

e.g., results/lstm.dataset.batchsize-16.epochs-500.opt-adam.log

As Things Progress...



Challenges surface

GPU sharing and infrastructure management

Experiment tracking for reproducibility and collaboration

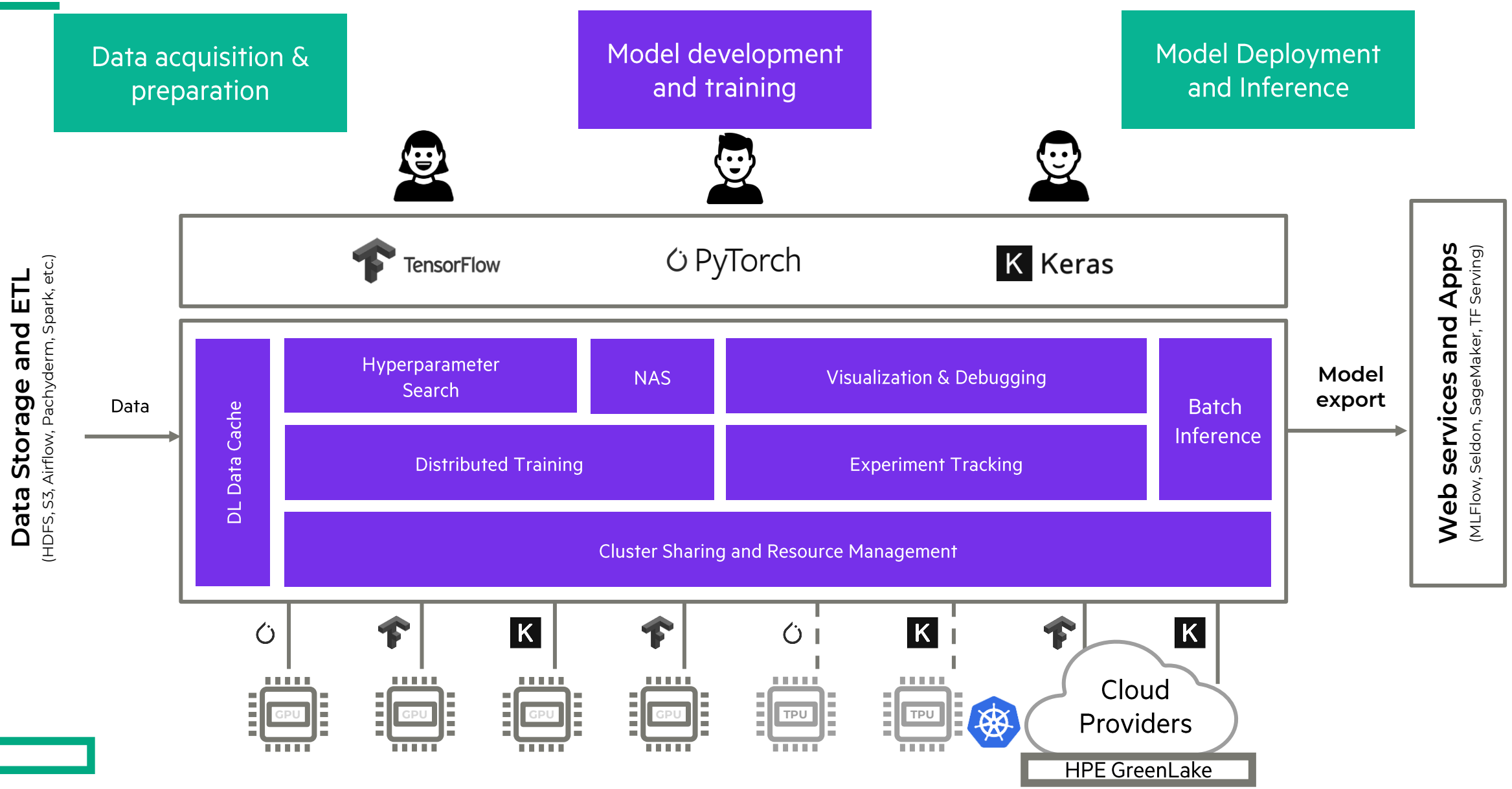
Hyperparameter Optimization

Ad-hoc, time-consuming, often no tool support

Distributed Training

Hard to configure, fragile, not multi-user by default

HPE Machine Learning Development Environment built on Determined AI



Top Tech Companies

Invested massive resources and expertise in custom-built internal DL software infrastructure.

Results:

- Infra itself as a comparative advantage
- Efficient, productive DL teams
- Transformative DL-powered applications

Everybody Else

Lack resources and expertise to build DL software infrastructure.

Results:

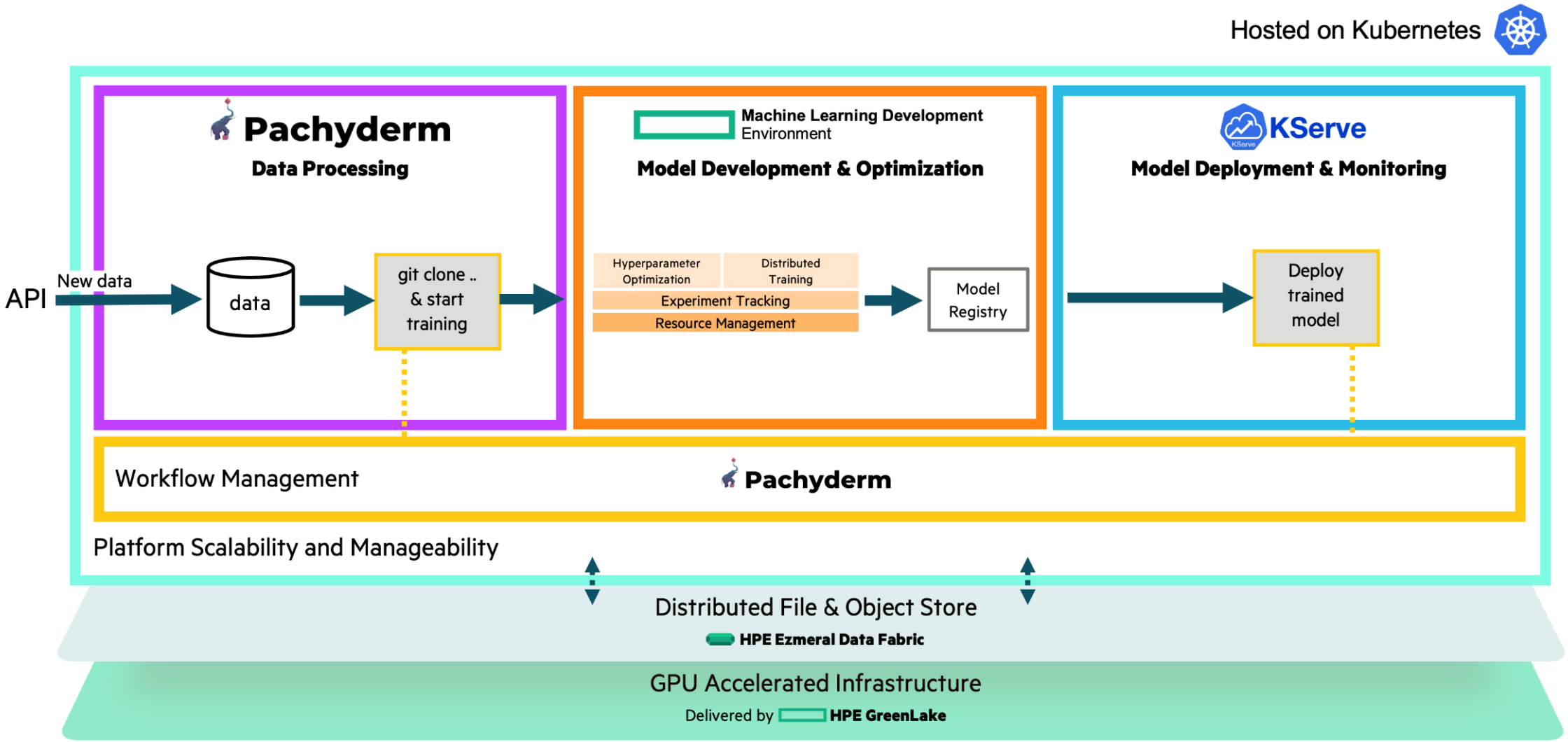
- Reliance on narrow, single-user tools (e.g., TF, Horovod)
- Frustrated, unproductive DL teams
- DL projects are wildly expensive, time-consuming, and frequently unsuccessful

We provide world-class DL software infrastructure for everyone

The problem: The AI landscape is growing and there is lack of common standards and cross-platform interoperability

The Solution: Open Stacks

Early Innings of ML

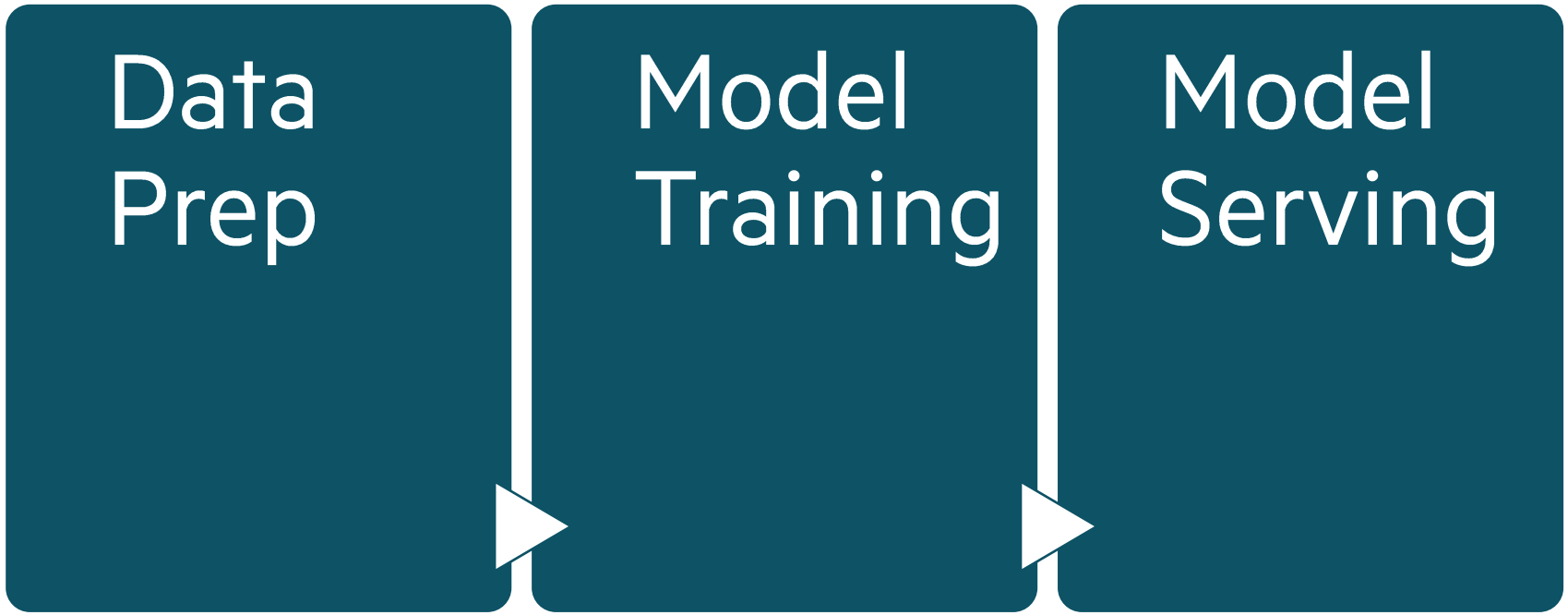


About AI Infrastructure Alliance (AIIA)

Establishing standards and canonical stack in AI/ML development

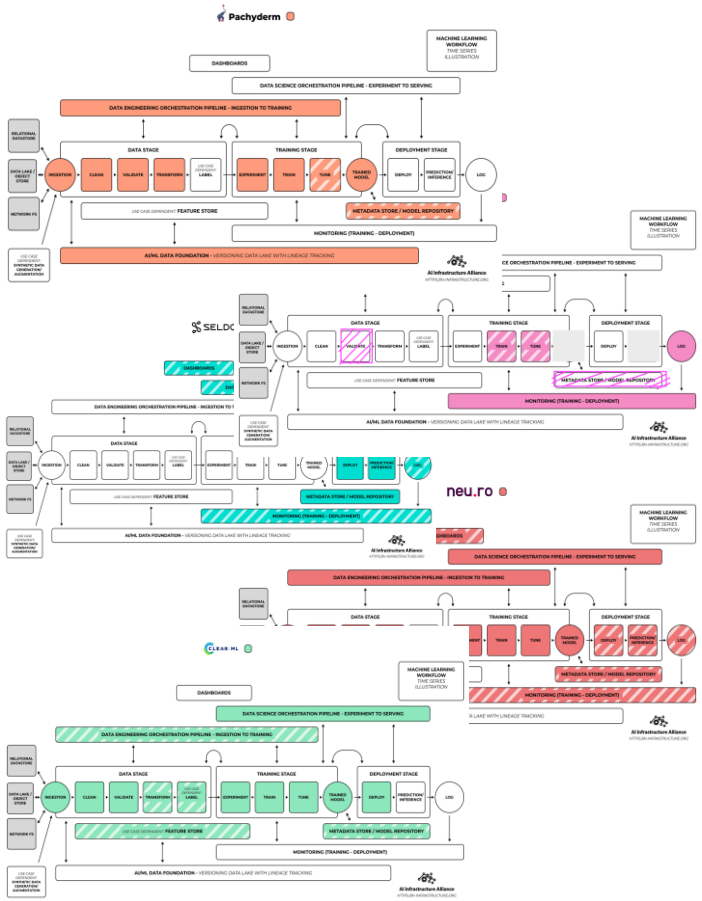
- Founded in late 2020
- Full launch in Feb 2021
- 501(c)(6) Delaware org since Oct 2021

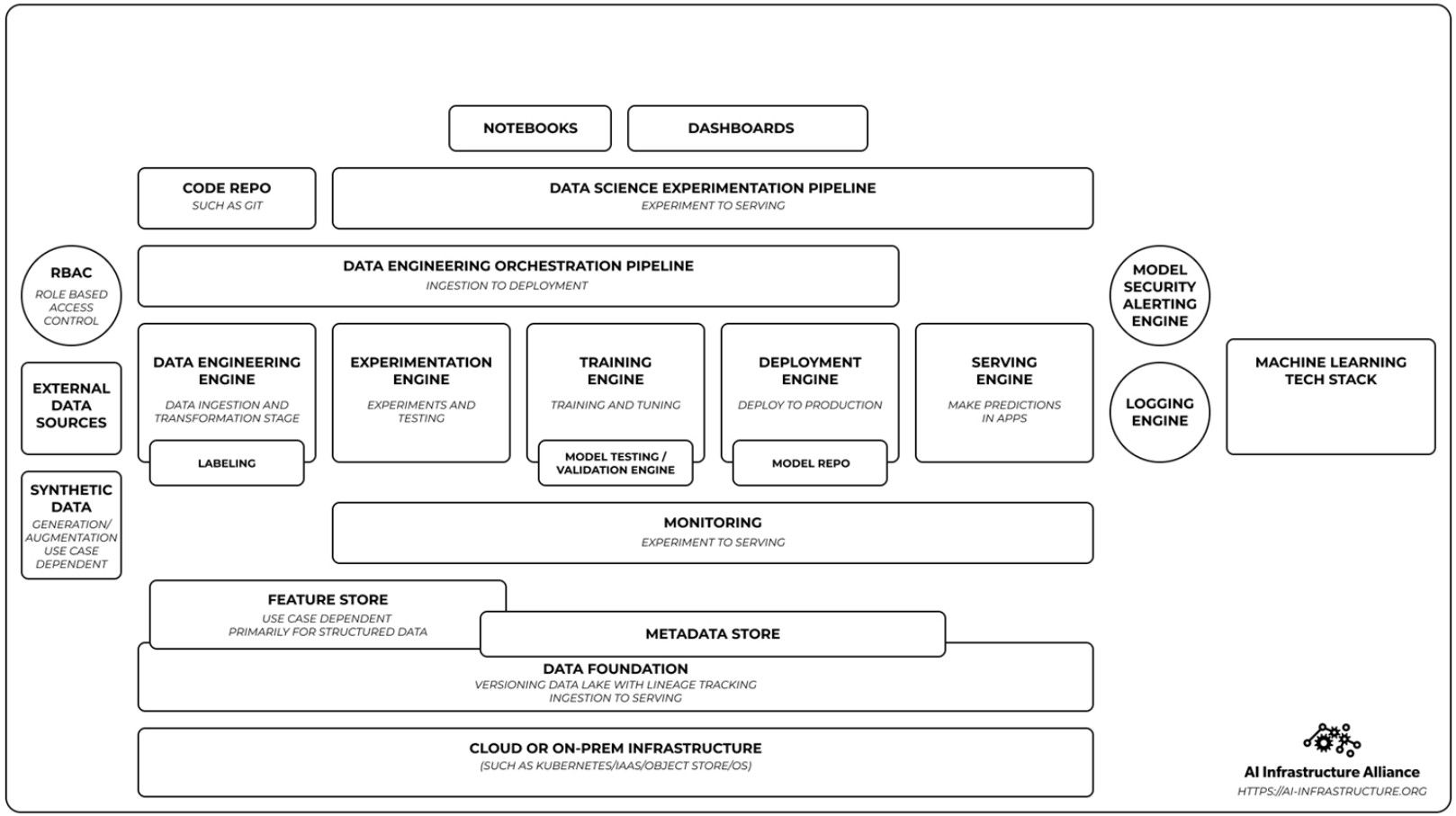
The Model Development Lifecycle (simplified)



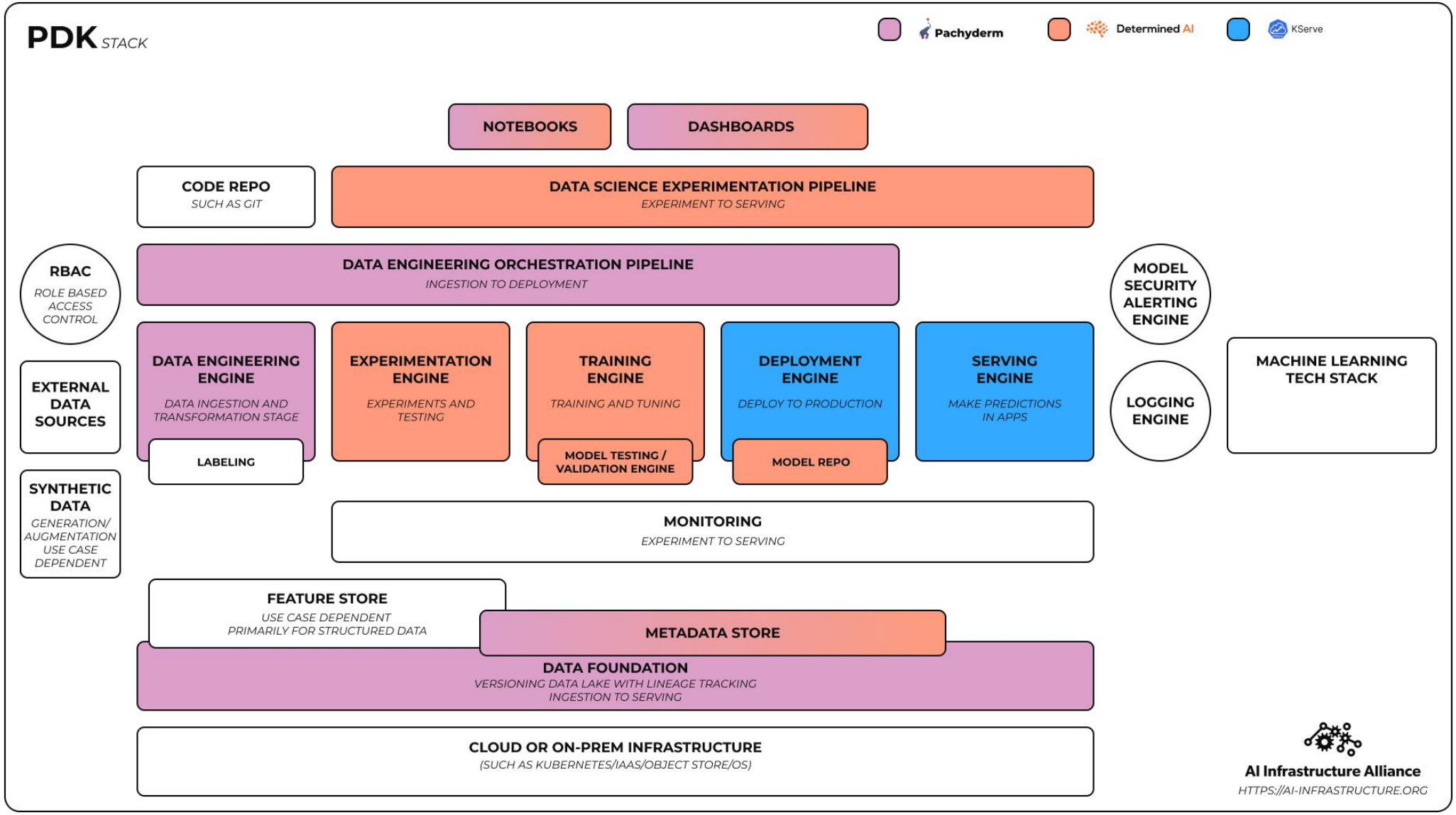
Blueprints widely adopted / Industry-wide push for open standards

Our blueprints were widely adopted across the industry with multiple companies having their own complete stack.



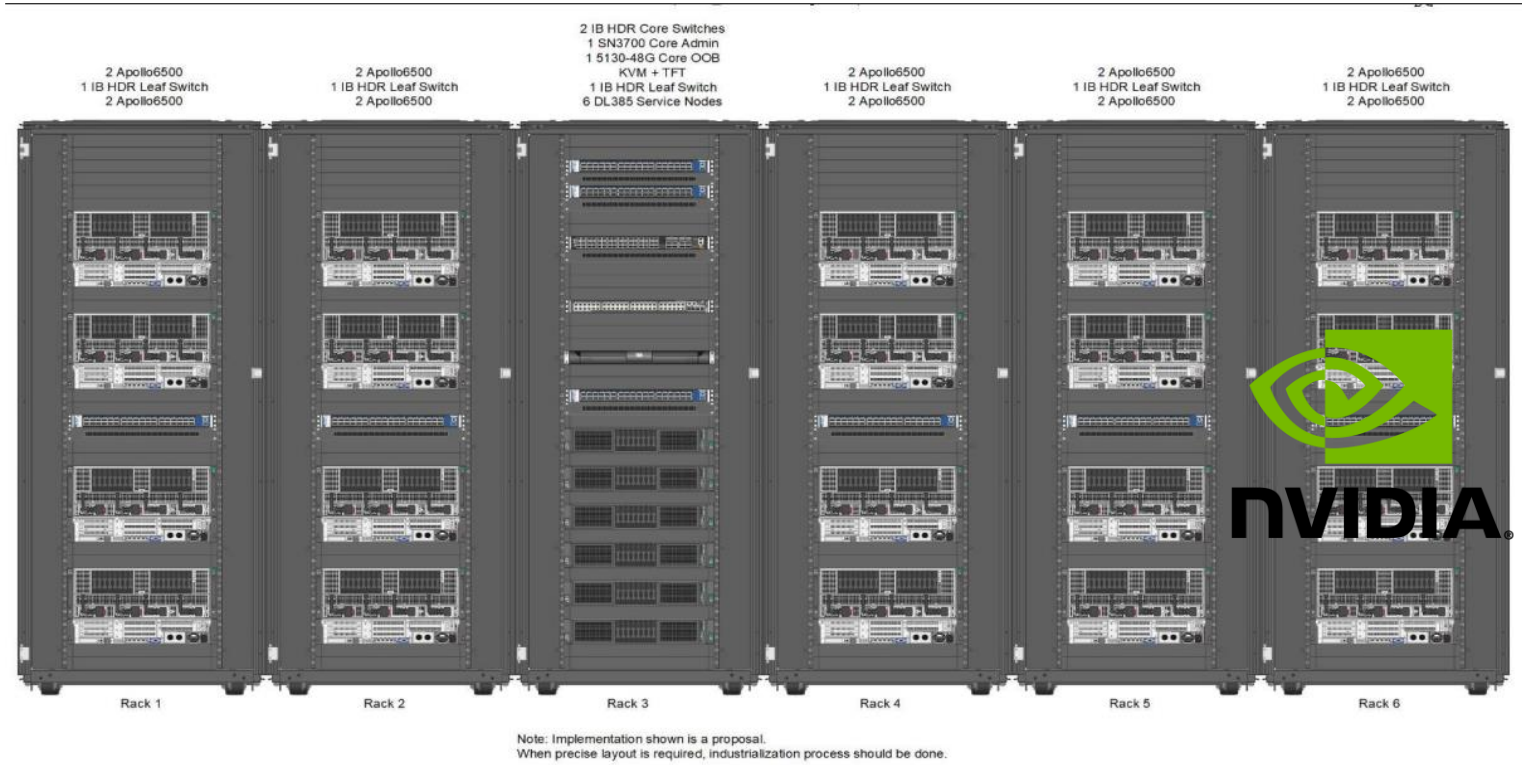


Interactive Blueprints



HPE/NVIDIA “Champollion” Supercomputer
New AI Supercomputer Accelerates Customers’ AI Journey

Hewlett Packard Enterprise



- New 20-node supercomputer hosted in HPE Center of Excellence (CoE) Grenoble, France to drive advanced testing and benchmarking
- Built in collaboration with NVIDIA, the supercomputer features 20 HPE Apollo Gen10 Plus systems, 160 NVIDIA A100 Tensor Core GPUs, and Infiniband networking solutions to advance AI and target-data and image-intensive workloads.
- Supercomputer aimed to help the scientific and engineering communities accelerate their AI journey – in making meaningful impact to key industries such as healthcare, transportation, and environment.

Let’s take a closer look at Determined

- (PyTorch MNIST Tutorial)

Thank you

Email: isha.ghodgaonkar@hpe.com
Twitter: @ishabytes

Determined AI Community Slack

Determined AI Homepage