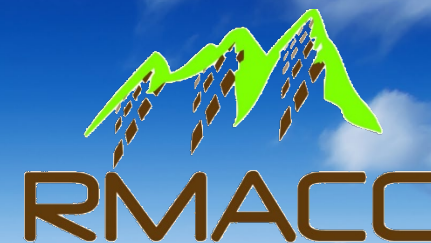




**Hewlett Packard
Enterprise**



Rocky Mountain Advanced Computing Consortium

FAST AND EASY WAY TO DEVELOP MACHINE LEARNING MODELS

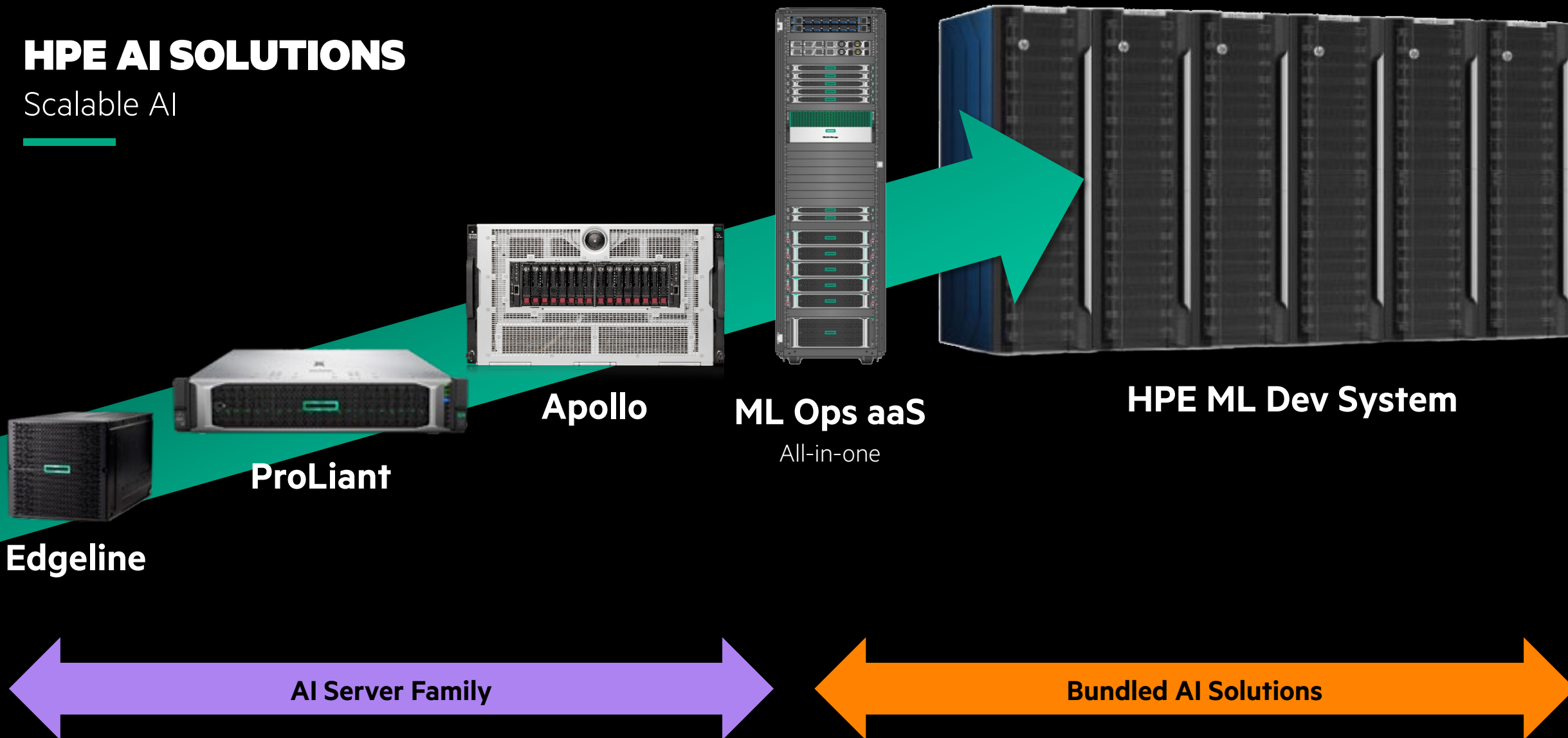
HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT

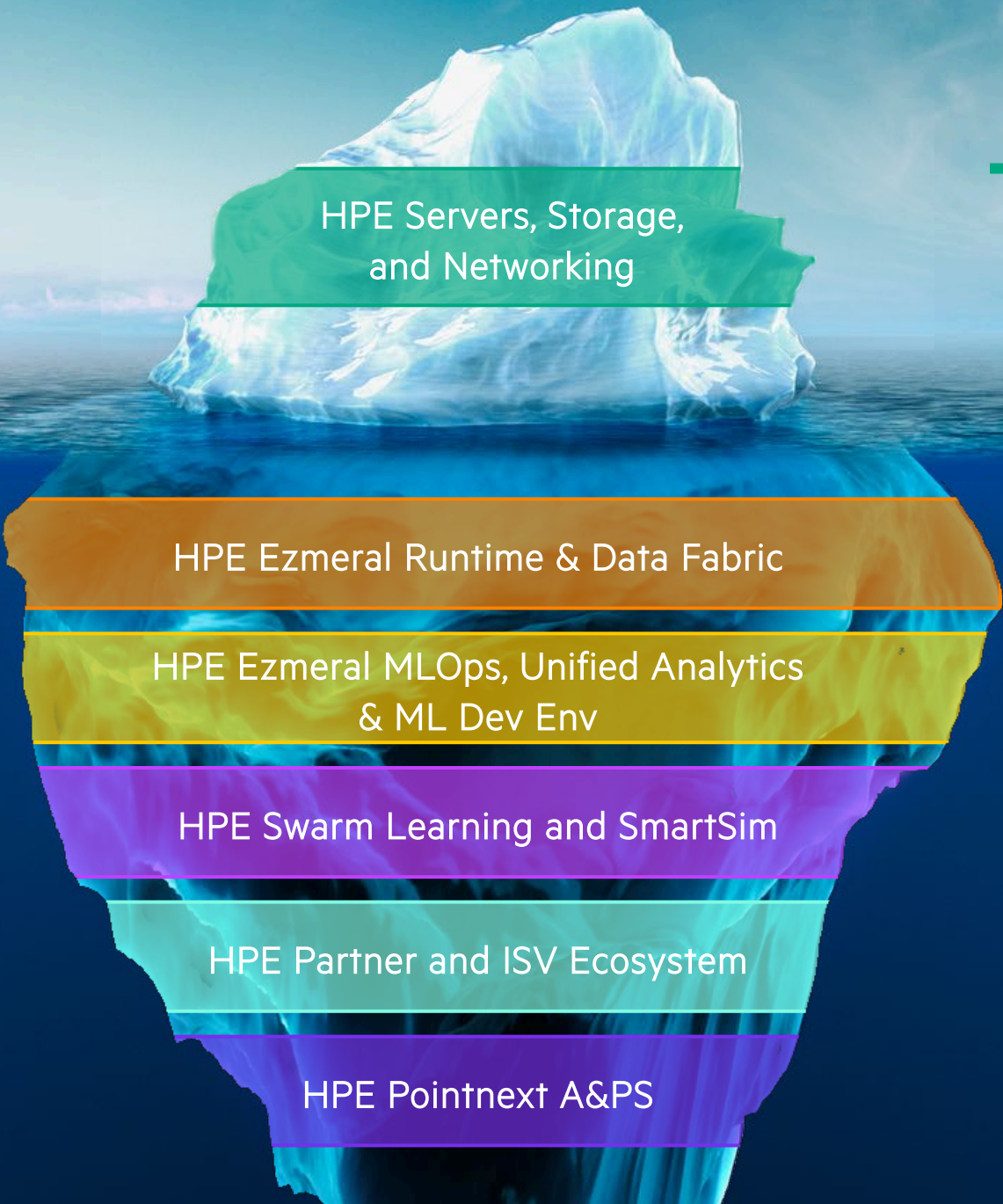
Steve Heibein
US Public Sector AI Lead

August 2022

HPE AI SOLUTIONS

Scalable AI



An iceberg floating in a blue ocean under a cloudy sky. The iceberg is divided into six horizontal layers, each with a different color and text. The top layer is green and is above the water line. The other five layers are below the water line and are colored orange, yellow, purple, teal, and dark purple from top to bottom. A large green bracket on the right side of the iceberg groups the five submerged layers together.

HPE Servers, Storage,
and Networking

HPE Ezmeral Runtime & Data Fabric

HPE Ezmeral MLOps, Unified Analytics
& ML Dev Env

HPE Swarm Learning and SmartSim

HPE Partner and ISV Ecosystem

HPE Pointnext A&PS

Full Stack
Production Ready
AI Solutions
delivered aaS
by HPE GreenLake

MODEL DEVELOPMENT AND TRAINING CHALLENGES

Data acquisition
and preparation

Model development and training

Key challenges

Training deep
learning models
is **complex**

Specialized infrastructure
needed to deliver complex
AI/ML workloads

ML engineers are focused on
managing infrastructure,
not delivering value

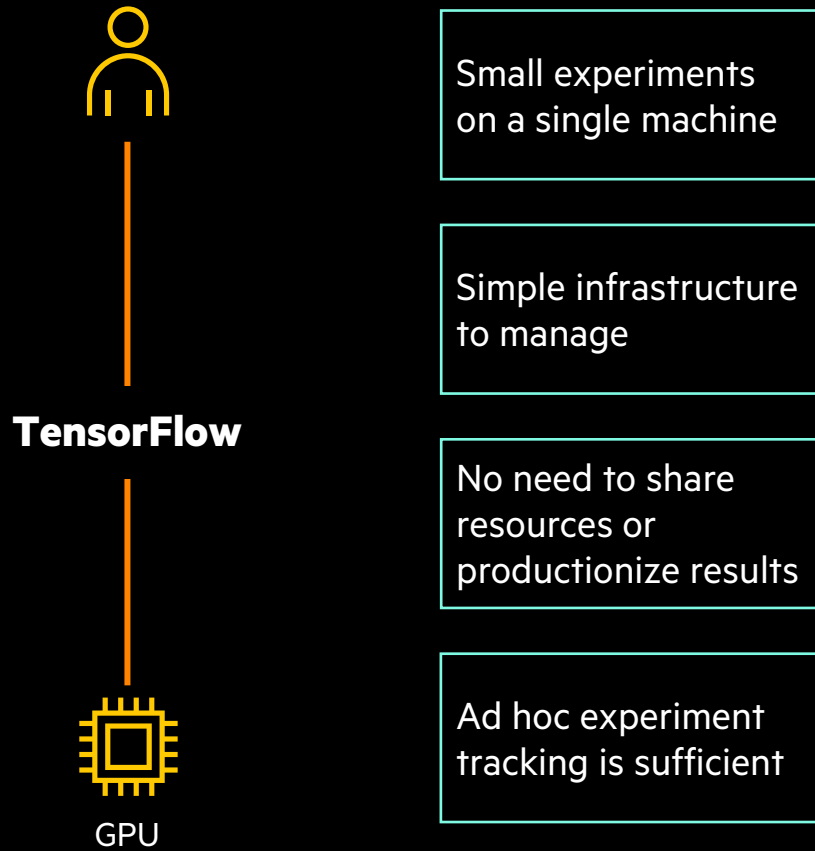
Cloud vendors and
specialized hardware
vendors are **locking you in**

Deployment
and inference

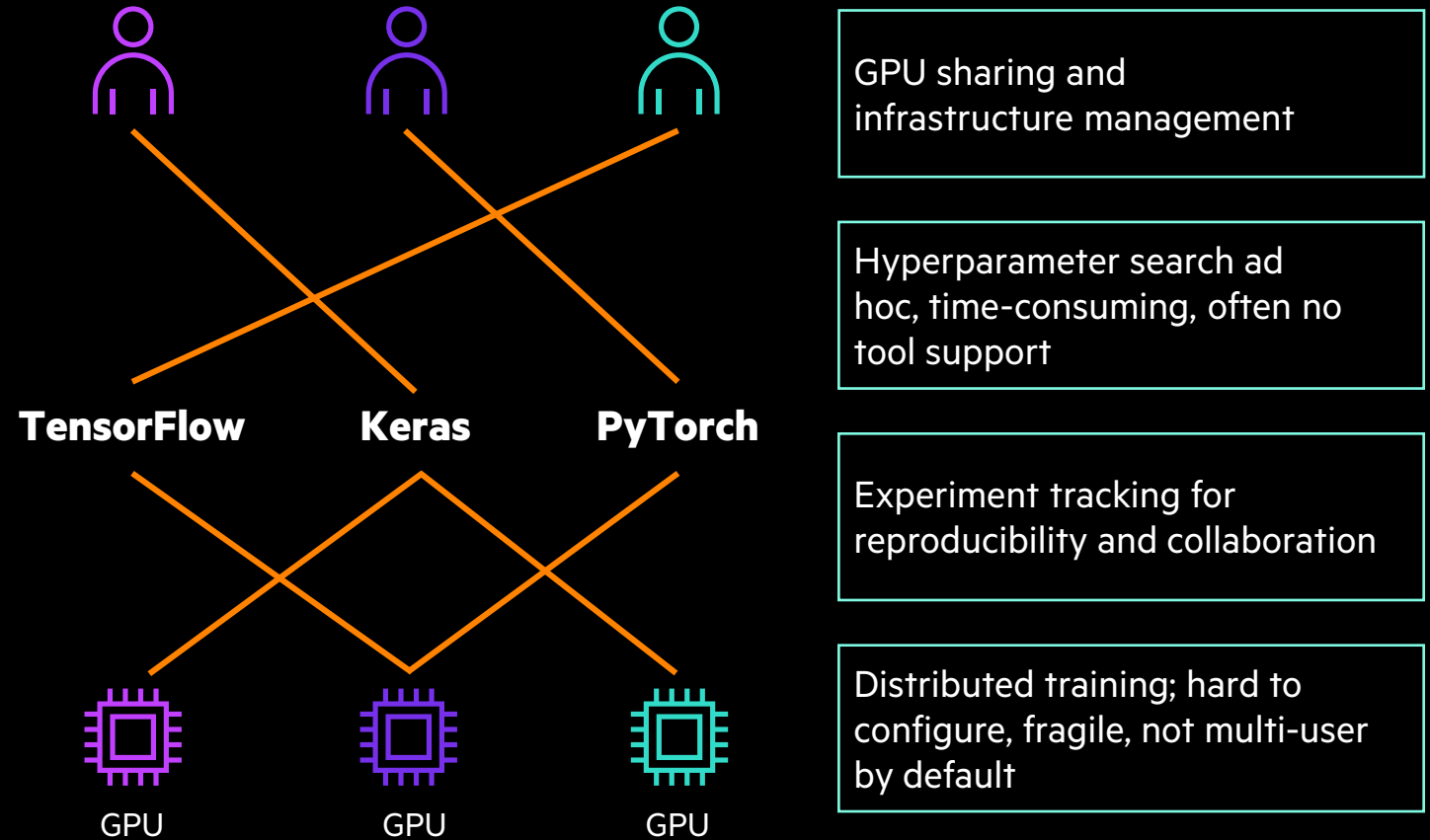
Standard compute infrastructure

HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT

On day one...small and simple



On day two....complex and complicated



BEST-IN-CLASS ML TOOLING

Buy vs. Build

The elite few

Only a limited number of companies can invest the massive resources and expertise needed to custom-build internal ML software infrastructure.

Results:

- Infrastructure itself becomes a business advantage
- Efficient, productive ML teams
- Transformative ML-powered applications

The many

Most companies lack the resources and expertise to build competitive ML software infrastructure.

Results:

- Reliance on narrow, single-user, open-source tools
- Frustrated, unproductive ML teams
- ML projects are wildly expensive, time-consuming, and frequently unsuccessful

The answer: Make world-class ML software infrastructure available to anyone.

HPE Machine Learning Development Environment enables you to access the same AI technology previously available only to larger companies like Facebook and Google **at a fraction of the cost and time.**

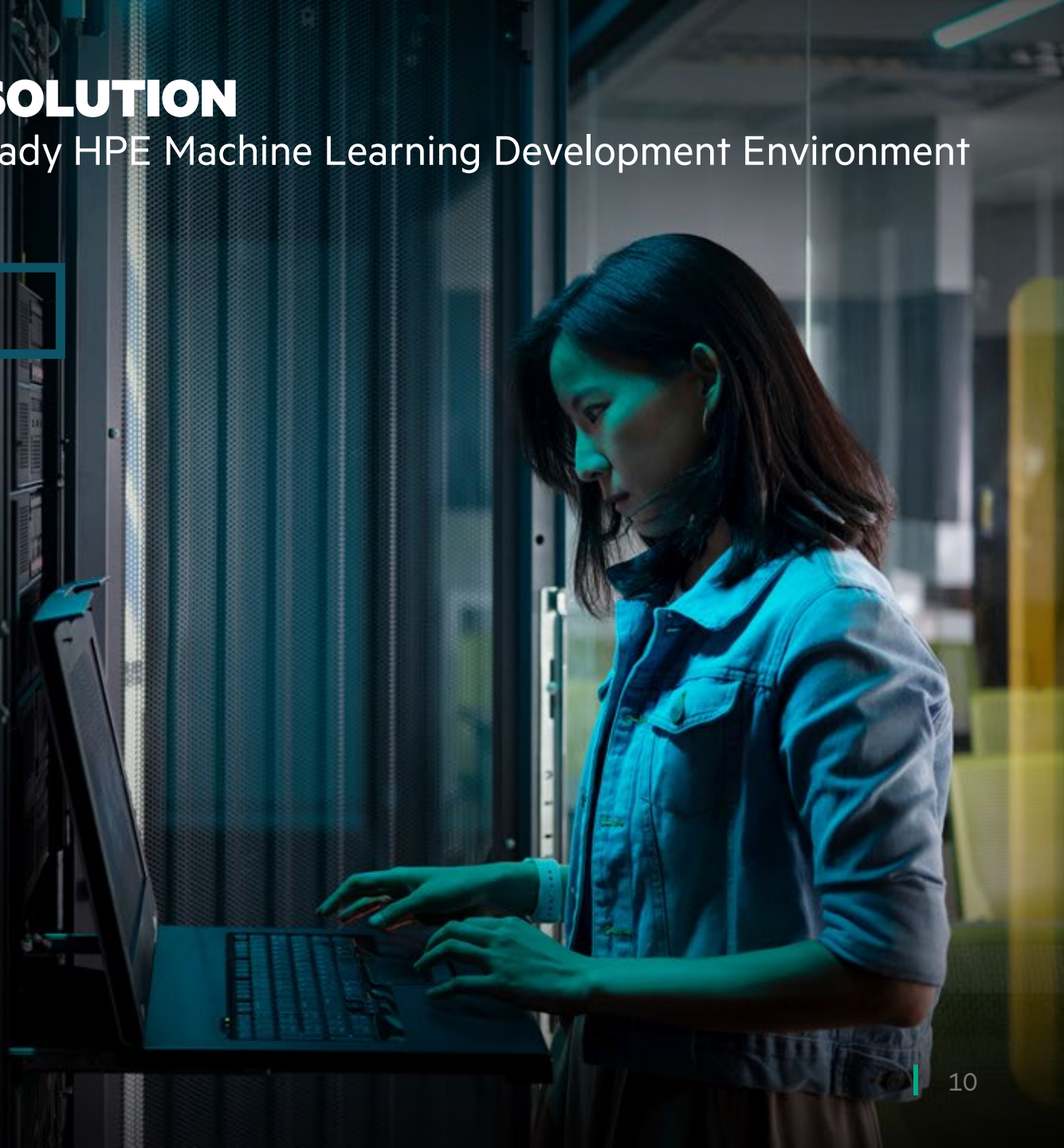


EVOLVING FROM AN OPEN-SOURCE SOLUTION

Open-source Determined AI becomes enterprise-ready HPE Machine Learning Development Environment

Determined AI | **GitHub** | **Slack**

- Providing the software infrastructure to make DL accessible to anyone
- Making Determined AI enterprise-ready
- Now part of HPE
- Enabling you to use, contribute, and provide feedback on Determined AI via:
 - [GitHub](#)
 - [Slack](#)

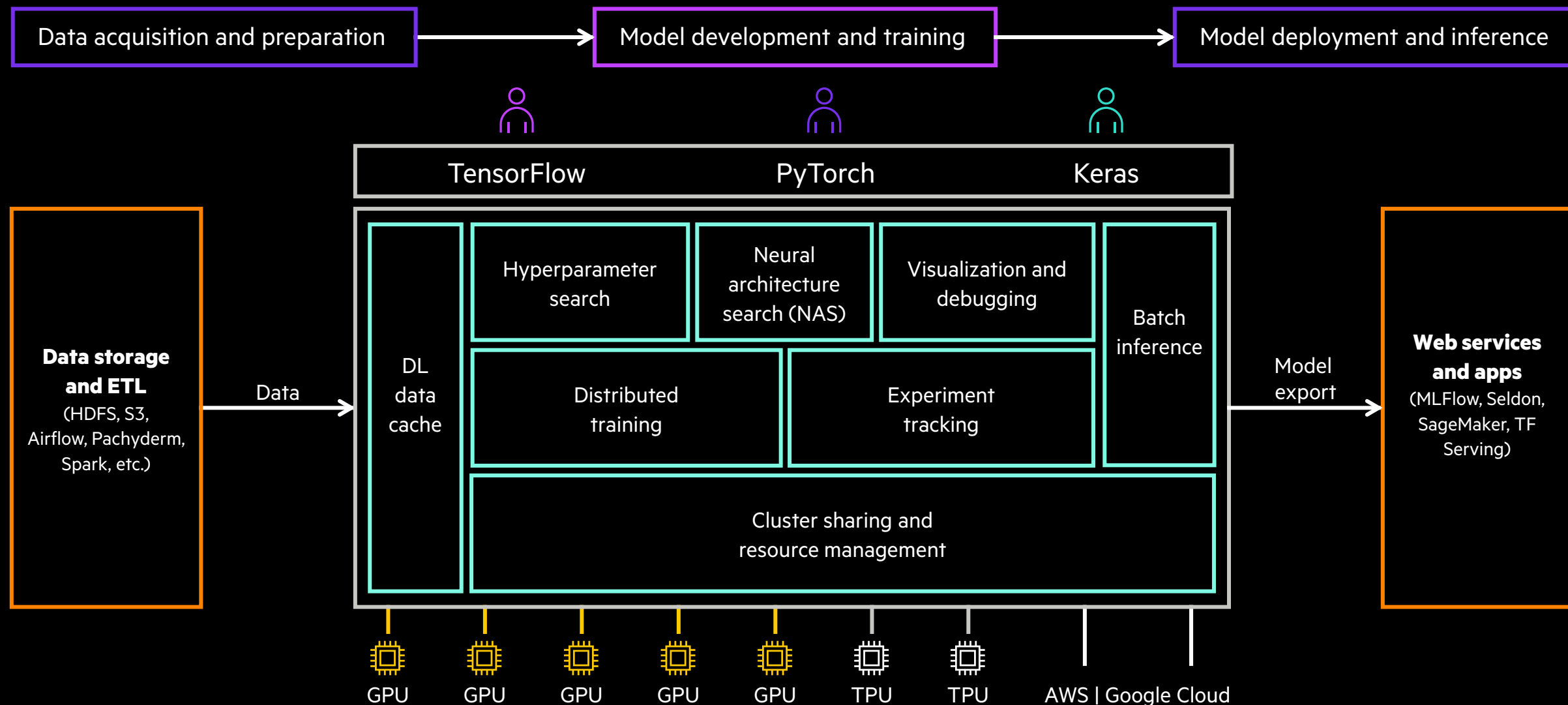


HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT AND DETERMINED OPEN SOURCE—COMPARED

	Open source Determined Software	HPE Machine Learning Development Environment
Distributed training	✓	✓
Model optimization	✓	✓
Metadata tracking	✓	✓
Cluster resource management	✓	✓
GPU cost management	✓	✓
Collaboration and experiment tracking	✓	✓
Security		
• Single sign on (SSO)	X	✓
• Automated user provisioning	X	✓
Premium dedicated support	X	✓



HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT




PRODUCT SCREENSHOTS

Determined AI
HPE Machine Learning Environment



RESOURCE MANAGEMENT



Determined AI

0.16.2

Launch JupyterLab

Dashboard

Experiments

Tasks

Cluster12.5%

Master Logs

Docs

API (Beta)

Collapse

determined

Cluster

OverviewHistorical Usage

Overview Stats

Connected Agents1


GPU Slots Allocated1 / 8

Overall Allocation

Compute (GPU) Slots Allocated

1 (13%)RUNNING0 (0%)PENDING7 (88%)FREE

3 Resource Pools



aux-pool

AWSDEFAULT AUX POOL

No Connected Agent

No description provided.

Compute Slots Allocated0 / 0

Aux containers running:0

Location:us-west-2

Instance Type:t2.xlarge

Spot/Preemptible:false

Min Agents:0


Max Agents:5

Slots Per Agent:0

Max Aux Containers Per Agent:100

Scheduler Type:Fairshare

View more info



compute-pool

AWSDEFAULT COMPUTE POOL

1 Connected Agent

No description provided.

GPU Slots Allocated1 / 8 (13%)

Aux containers running:0

Location:us-west-2

Instance Type:p2.8xlarge

Spot/Preemptible:false

Min Agents:0


Max Agents:5

Slots Per Agent:8

Max Aux Containers Per Agent:0

Scheduler Type:Fairshare

View more info



compute-pool-spot

AWS

No Connected Agent

No description provided.

GPU Slots Allocated0 / 0

Aux containers running:0

Location:us-west-2

Instance Type:p2.8xlarge

Spot/Preemptible:true

Min Agents:0

Max Agents:5

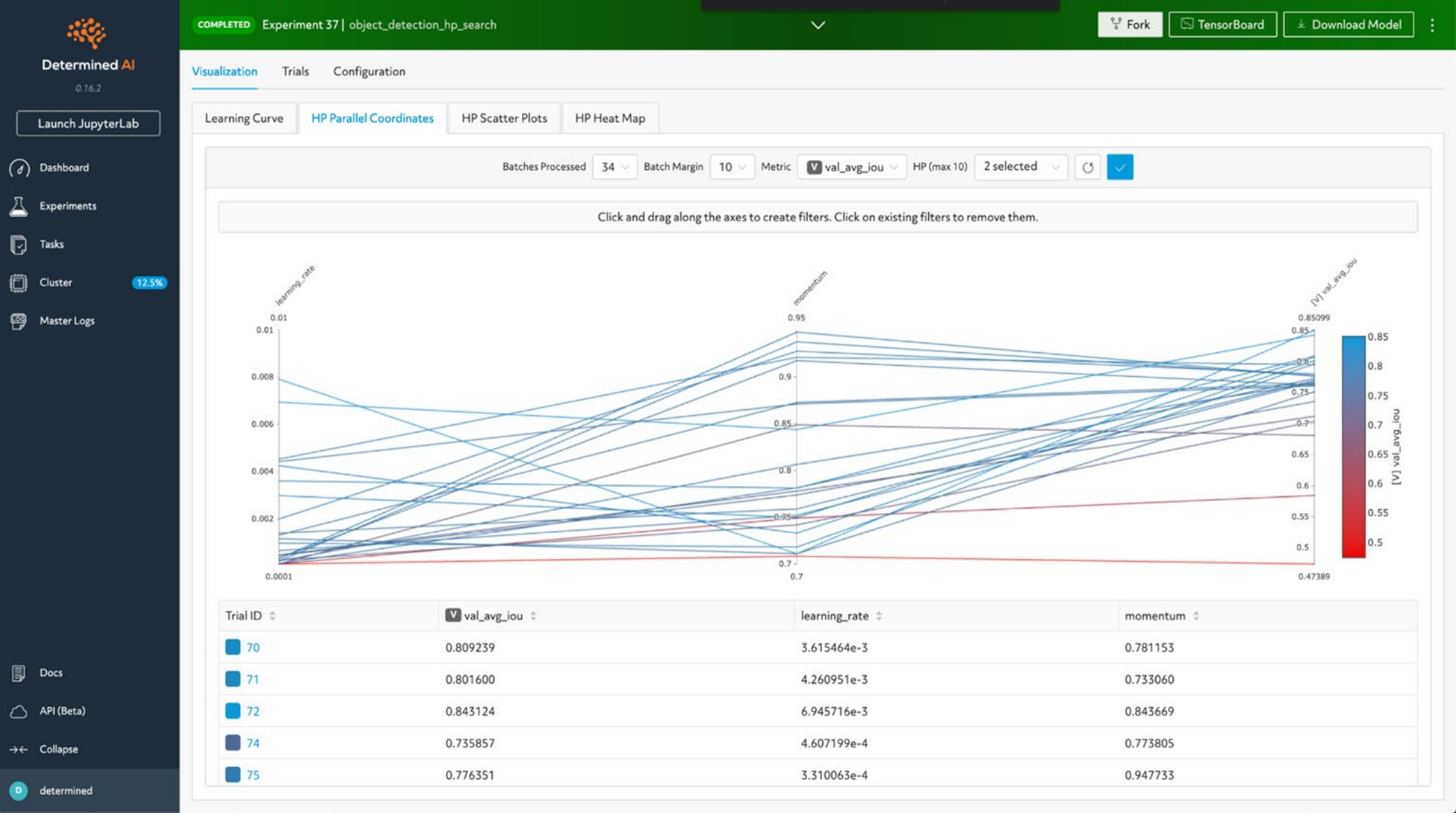
Slots Per Agent:8

Max Aux Containers Per Agent:0

Scheduler Type:Fairshare

View more info

HYPERPARAMETER OPTIMIZATION



NOTEBOOK ENVIRONMENT

Filter files by name


Name	Last Modified
Y: const.yaml	7 days ago
data.py	8 months ago
Y: distributed...	7 days ago
model_def....	7 days ago
models.py	8 months ago
Y: notebook....	7 days ago
Object-De...	7 days ago
predict.py	8 months ago
ptl.py	8 months ago
Y: search.yaml	7 days ago
test.ipynb	4 years ago
test.jpg	8 months ago

Launcher

Object-Detection-PyTorch.

Python 3

Building a Pedestrian Detection Model with Determined




This notebook will walk through the benefits of building a Deep Learning model with Determined. We will build an object detection model trained on the [Penn-Fudan Database for Pedestrian Detection and Segmentation](#).

Table of Contents

1. What Modeling looks like Today
2. Building a model with Determined
 - A. Single GPU training
 - B. Cluster-scale multi-GPU training
 - C. Adaptive hyperparameter search

What modeling looks like without Determined



DRIVING GREATER VALUE

Boost speed and performance

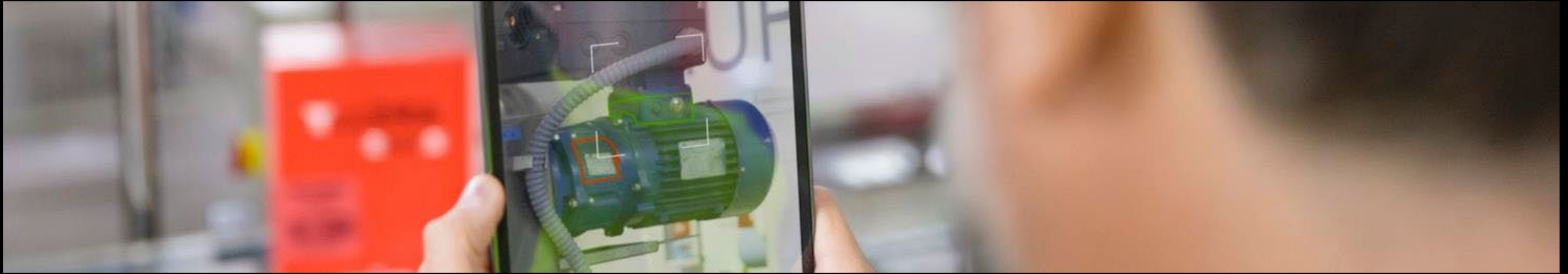
- **100x** faster than standard (random, grid)
- **10x** faster than research methods²
- Maximizing a model's predictive performance

Seamlessly scale to multiple machines

- **44x** faster training than on a single GPU¹
- **2x** faster distributed training³
- Up to **70%** savings on cloud instances¹
- Built-in provisioning support
- No infrastructure code needed

Simplify ML engineering

- Save each ML engineer **1 day's work per week**⁴
 - Using integrated ML tools
- Enable ML engineers to focus on models
 - **Reducing time to production**



Seamlessly integrated components —→ Dramatically easier to use

¹ Real-world computer vision application (Faster RCNN/COCO Object Detection).

² Based on peer-reviewed, published academic results.

³ Than Horovod at time of comparison via a gradient aggregation optimization

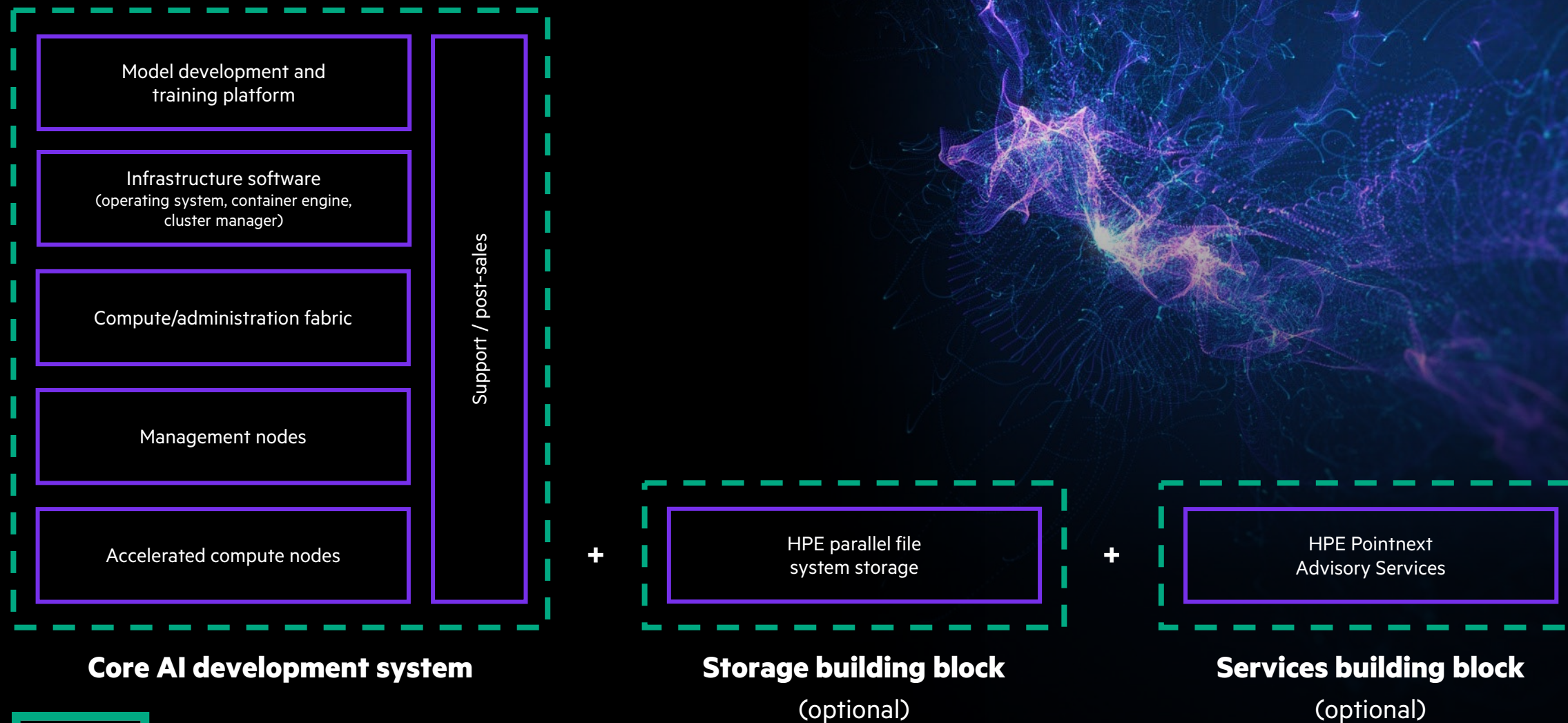
⁴ Based on customer feedback

HPE MACHINE LEARNING DEVELOPMENT SYSTEM



HPE MACHINE LEARNING DEVELOPMENT SYSTEM

Product packaging concept

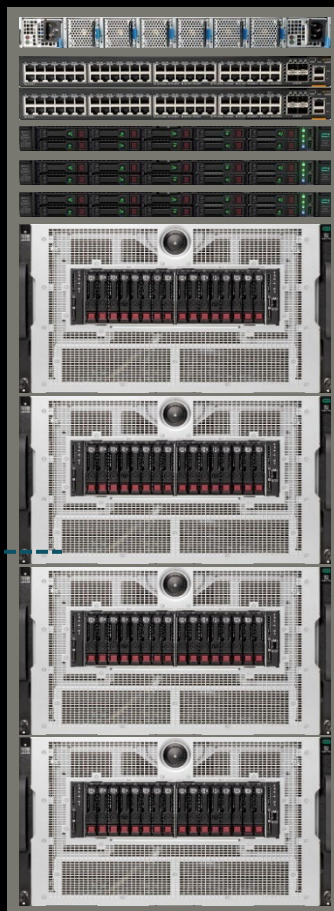


STANDARD HPE MACHINE LEARNING DEVELOPMENT SYSTEM OFFERING

Small system (4 nodes)

Infrastructure

- 1x IB HDR switch
- 2x Aruba 6300M Gbe switch
- 4 x HPE Parallel File System (Optional)
- 3 x HPE ProLiant DL325 Gen10+ Service Nodes
- 4 x HPE Apollo 6500 Gen10+ (8x 80GB NVIDIA A100 GPUs)



+

Services and software stack¹

Services and support

Pre-installed, configured and benchmarked in factory

On-site system start up and validation (including software)

“Getting Started” one-day workshop

Full solution support

HPE Pointnext Services

Software

Training platform

Container runtime

Cluster manager

Operating system

HPE Machine Learning Development Environment

Docker

HPE Performance Cluster Manager

Red Hat Enterprise Linux®

CPU options:

- 2x AMD EPYC 7763
- 2x AMD EPYC 7543

Options (per node):

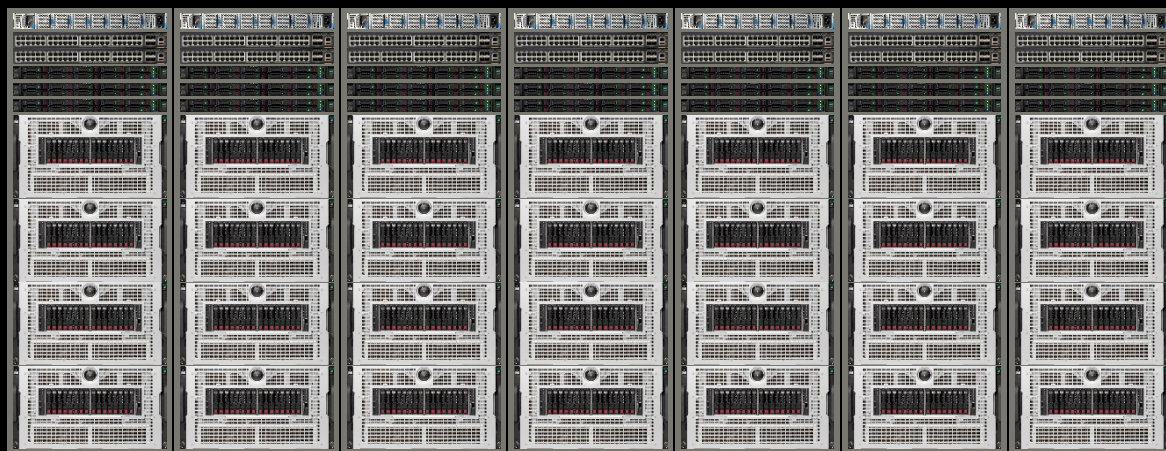
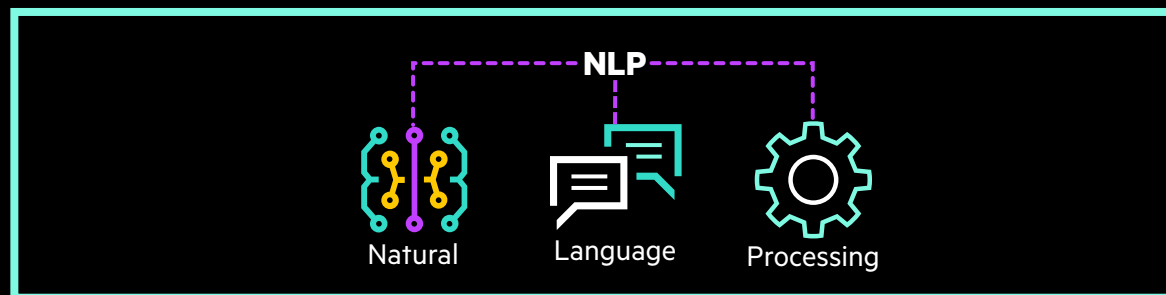
- Standard (2 TB RAM, 15 TB NVMe scratch)
- Performance (4 TB RAM, 30 TB NVMe scratch)

¹ Services design is WIP with HPE Pointnext Services

SOLUTION IN ACTION

HPE Machine Learning Development System delivers outcomes that matter

HPE Machine Learning Development System



Problem:

Need a new solution for developing large NLP models for both training and inference

Solution:

HPE Machine Learning Development System composed of:

- 64x HPE Apollo A6500 Gen10+ Systems
- 512 NVIDIA A100 GPUs
- HPE Parallel File System (GPFS)
- HPE Machine Learning Development Environment

Outcomes:

- Model parallelism
- Faster time to outcome/value
- Customized hyperparameter optimization
- Experiment tracking for collaboration
- Ability to develop on-premises and on the cloud

'CHAMP'OLLION

HPE – NVIDIA AI Cluster

- ❑ Compute - 20 Apollo 6500 gen10+
 - 8 GPU Nvidia A100 SXM4 80GB with NVswitch
 - 2 AMD EPYC™ 7763 (64 cores @ 2.45 – 3.5 GHz)
- ❑ Network - Mellanox (Infiniband) HDR fabric with SHARP™
 - 4 InfiniBand HDR (connect-6 200GB/s)
 - 1 Ethernet Mellanox (connectX-6 100GB/s)
- ❑ Storage (optionality available based on workload needs)
 - HPE PFSS 188TB
 - ClusterStor E1000
- ❑ Software Options
 - Machine Learning Development Environment
 - NVIDIA NGC Catalog
 - Ezmeral Suite
 - Open Source



Rank	System	Rmax (PFlop/s)	Rpeak (PFlop/s)
372	Champollion - Apollo 6500, AMD EPYC 7763 64C 2.45GHz, NVIDIA A100 SXM4 80 GB, Mellanox HDR Infiniband, HPE Hewlett Packard Enterprise France	2.02	2.52



THANK YOU

For more information, visit:

- hpe.com/info/machine-learning-development-environment
- github.com/determined-ai/determined
- determined.ai

