Hewlett Packard Enterprise

RMACC
Rocky Mountain Advanced Computing Consortium

FAST AND EASY WAY TO DEVELOP MACHINE LEARNING MODELS

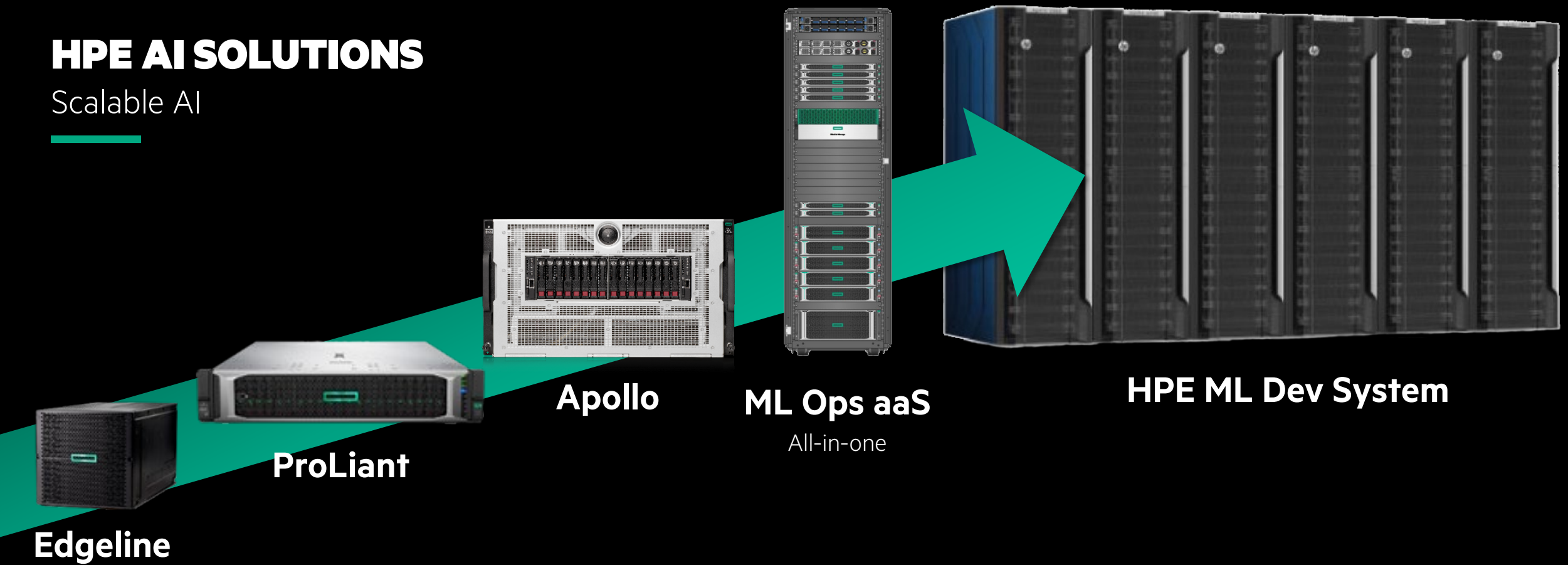
HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT

Steve Heibein
US Public Sector AI Lead

August 2022

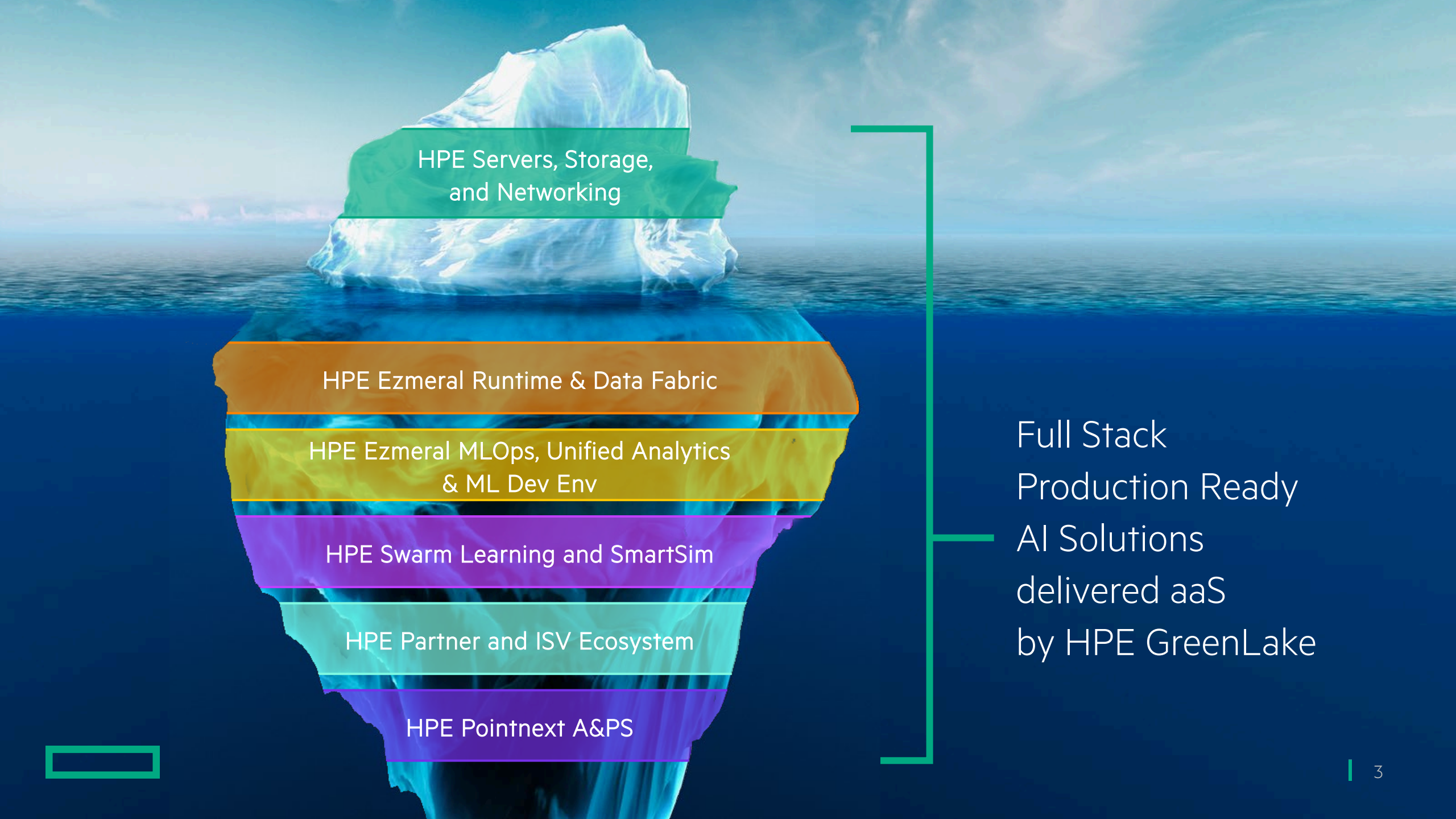
HPE AI SOLUTIONS

Scalable AI



AI Server Family

Bundled AI Solutions



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**

ML engineers are focused on **managing infrastructure**, not delivering value

Specialized infrastructure needed to deliver complex AI/ML workloads

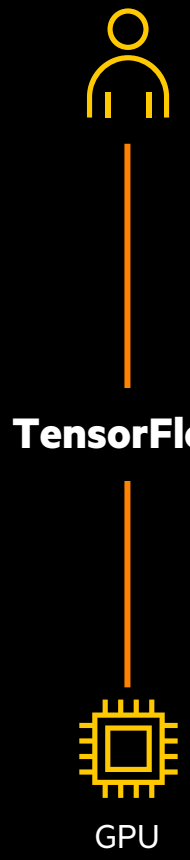
Cloud vendors and specialized hardware vendors are **locking you in**

Deployment and inference

Standard compute infrastructure

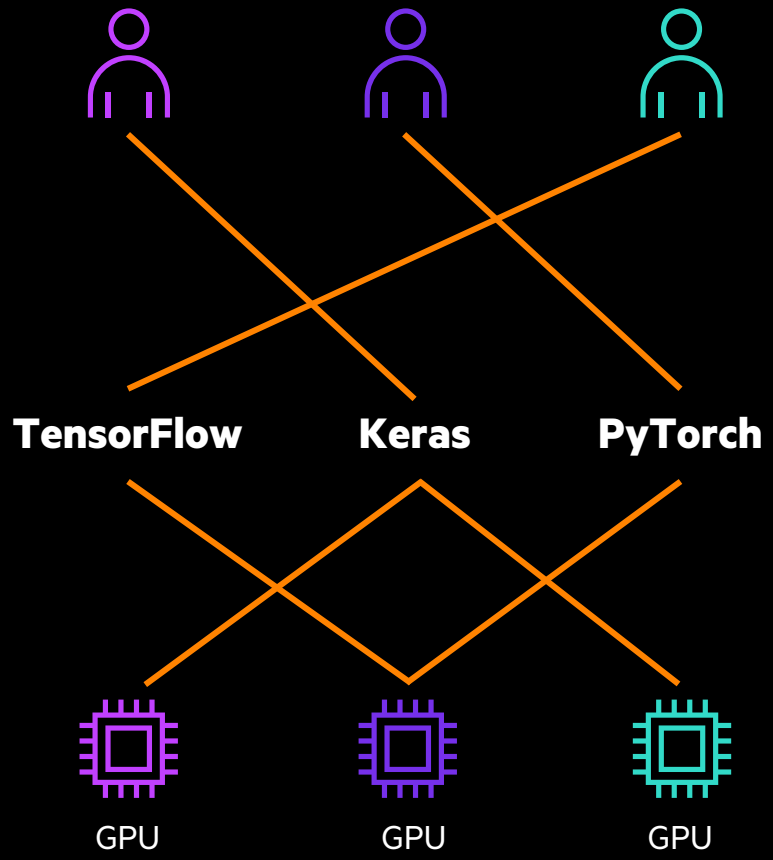
HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT

On day one…small and simple



- Small experiments on a single machine
- Simple infrastructure to manage
- No need to share resources or productionize results
- Ad hoc experiment tracking is sufficient

On day two….complex and complicated



- GPU sharing and infrastructure management
- Hyperparameter search ad hoc, time-consuming, often no tool support
- Experiment tracking for reproducibility and collaboration
- Distributed training; hard to configure, fragile, not multi-user by default

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.

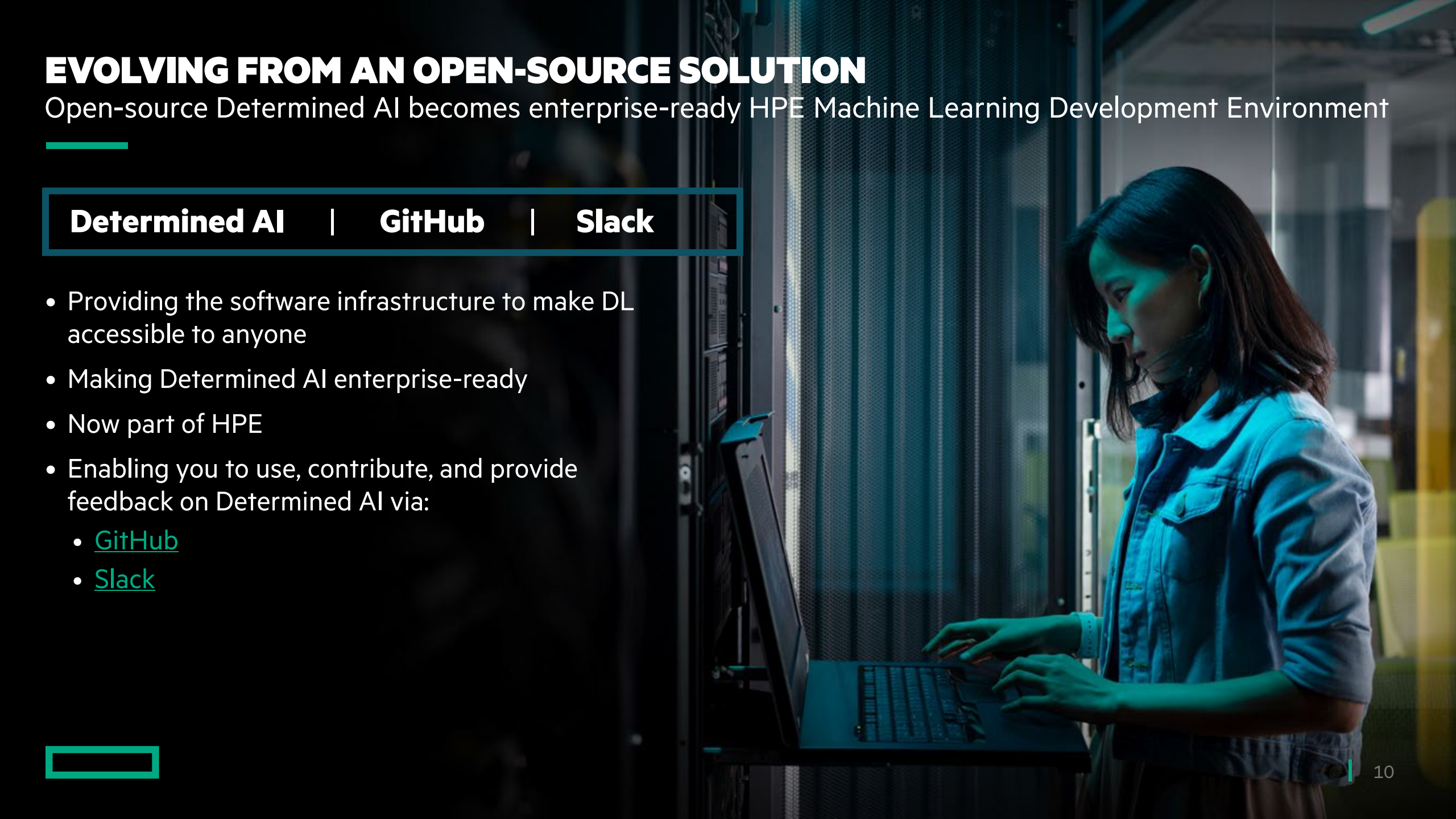
HPE Machine Learning Development Environment enables you to access the same AI technology previously available only to larger companies like Facebook and Google **at a fraction of the cost and time**.

EVOLVING FROM AN OPEN-SOURCE SOLUTION

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

Determined AI | GitHub | Slack

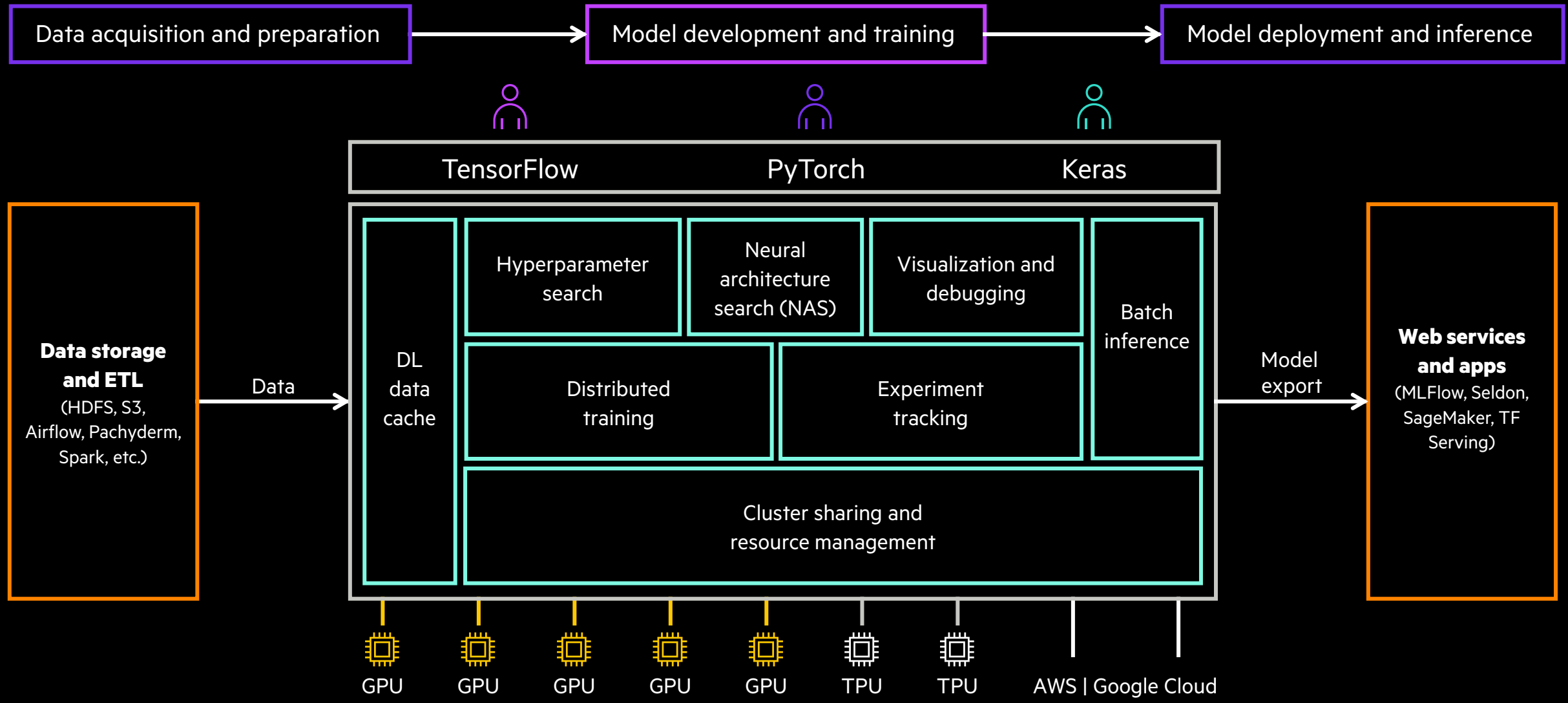
- Providing the software infrastructure to make DL accessible to anyone
- Making Determined AI enterprise-ready
- Now part of HPE
- Enabling you to use, contribute, and provide feedback on Determined AI via:
  - GitHub
  - Slack



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	✓

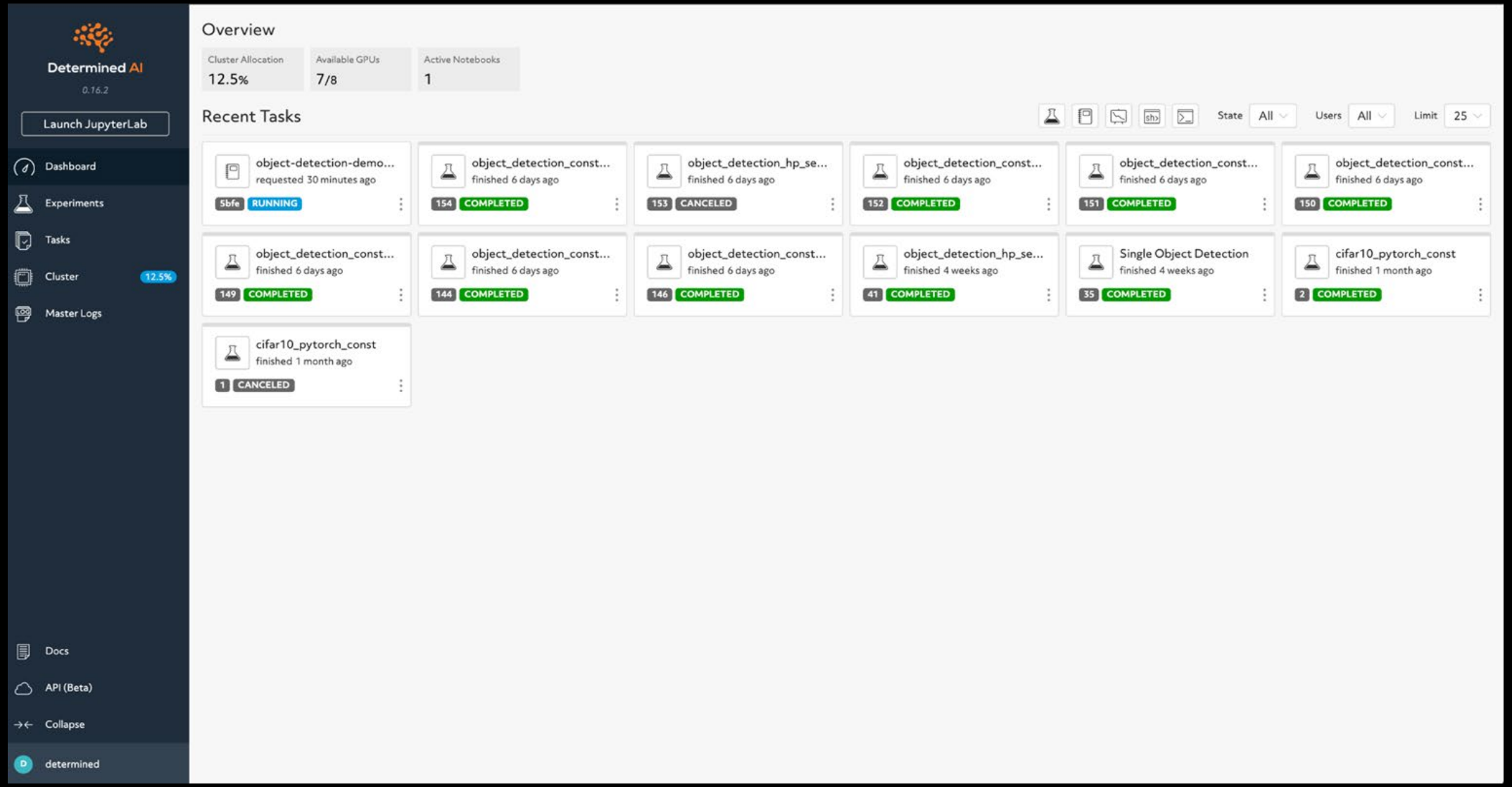
HPE MACHINE LEARNING DEVELOPMENT ENVIRONMENT



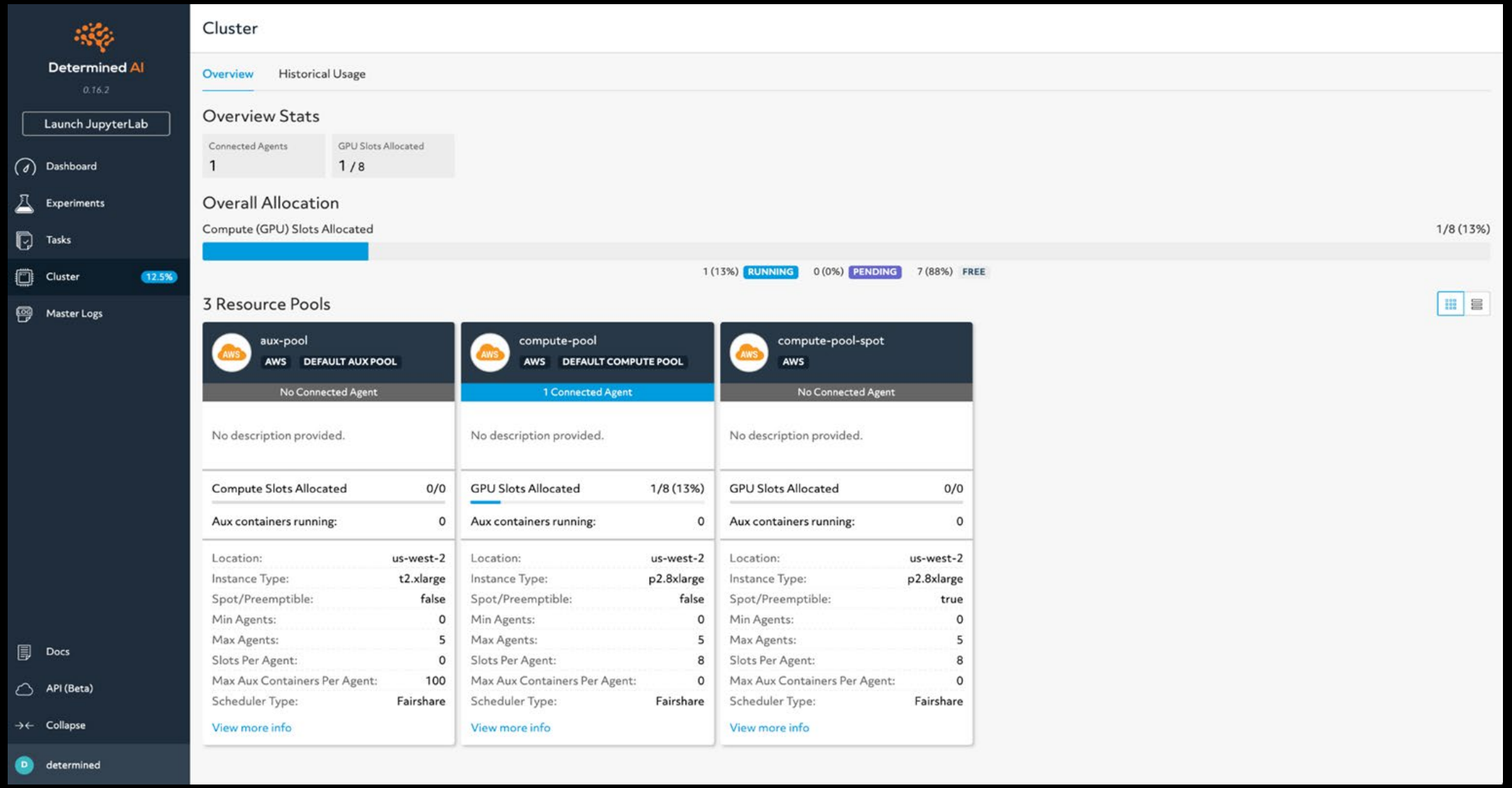
PRODUCT SCREENSHOTS

Determined AI
HPE Machine Learning Environment

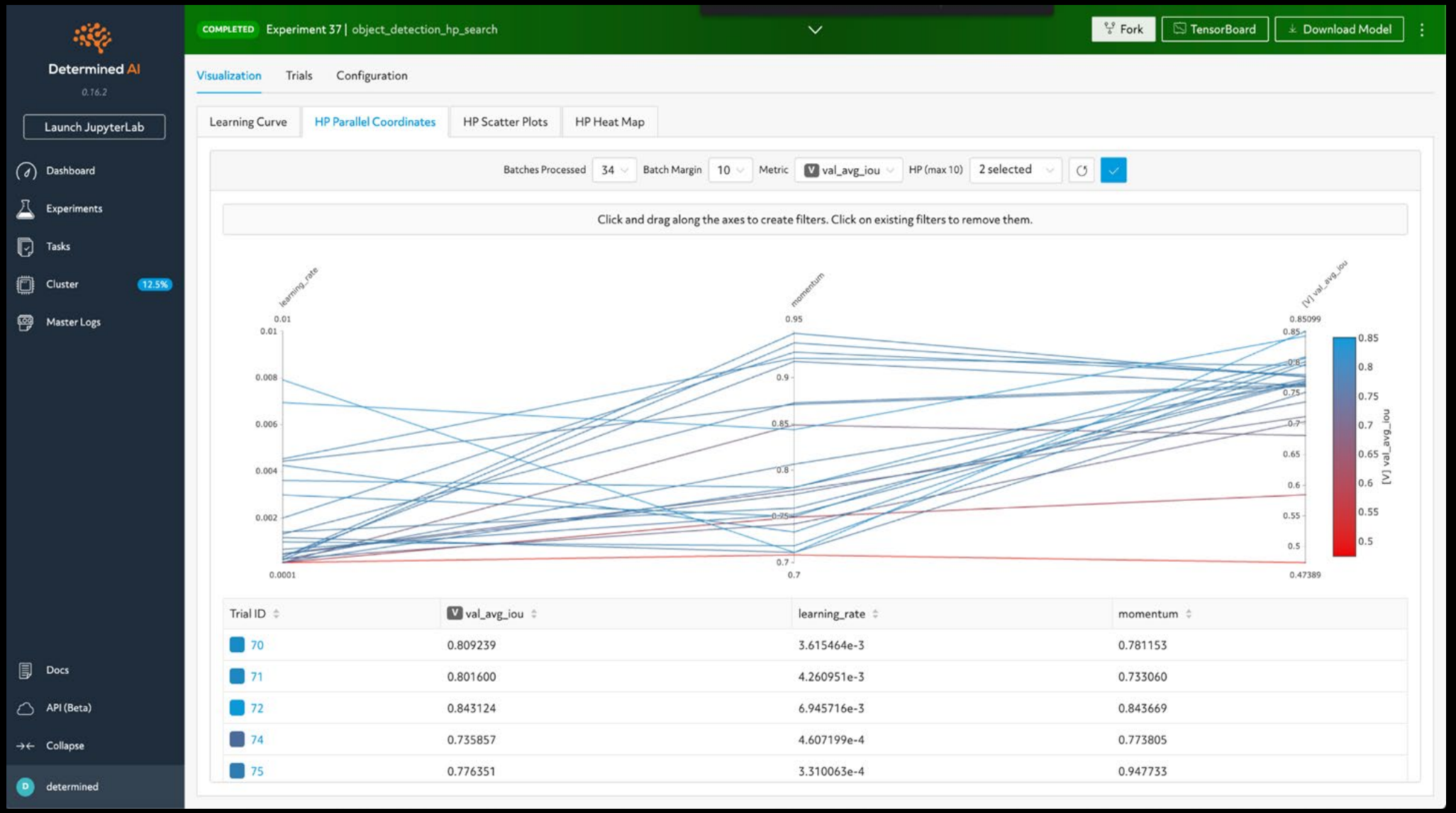
EXPERIMENT TRACKING



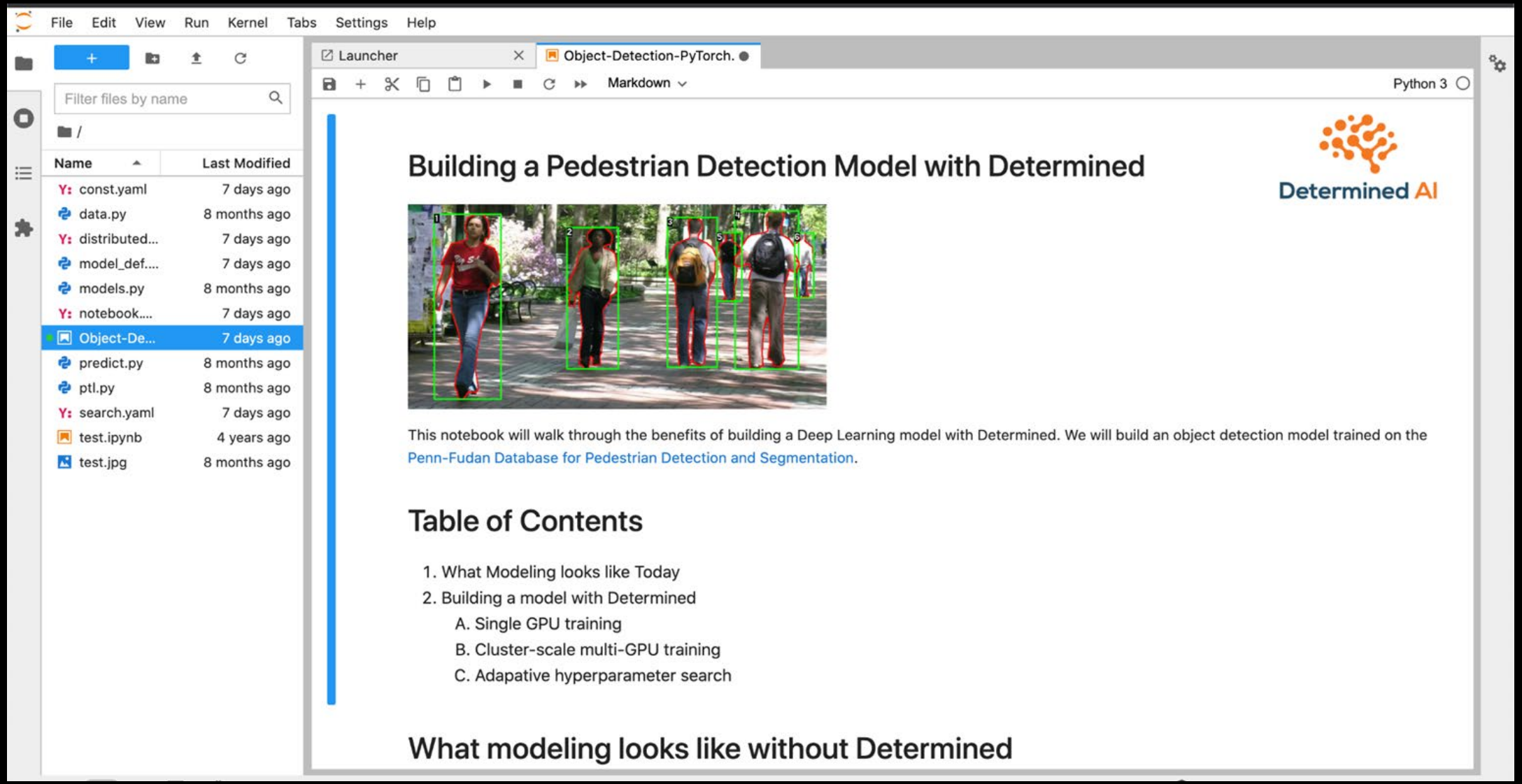
RESOURCE MANAGEMENT



HYPERPARAMETER OPTIMIZATION



NOTEBOOK ENVIRONMENT



DRIVING GREATER VALUE

Boost speed and performance

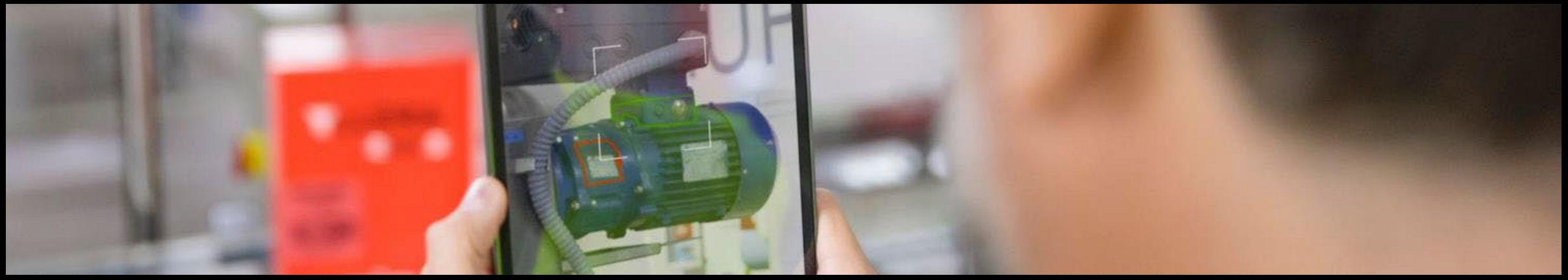
- **100x** faster than standard (random, grid)
- **10x** faster than research methods[2]
- Maximizing a model's predictive performance

Seamlessly scale to multiple machines

- **44x** faster training than on a single GPU[1]
- **2x** faster distributed training[3]
- Up to **70%** savings on cloud instances[1]
- 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**



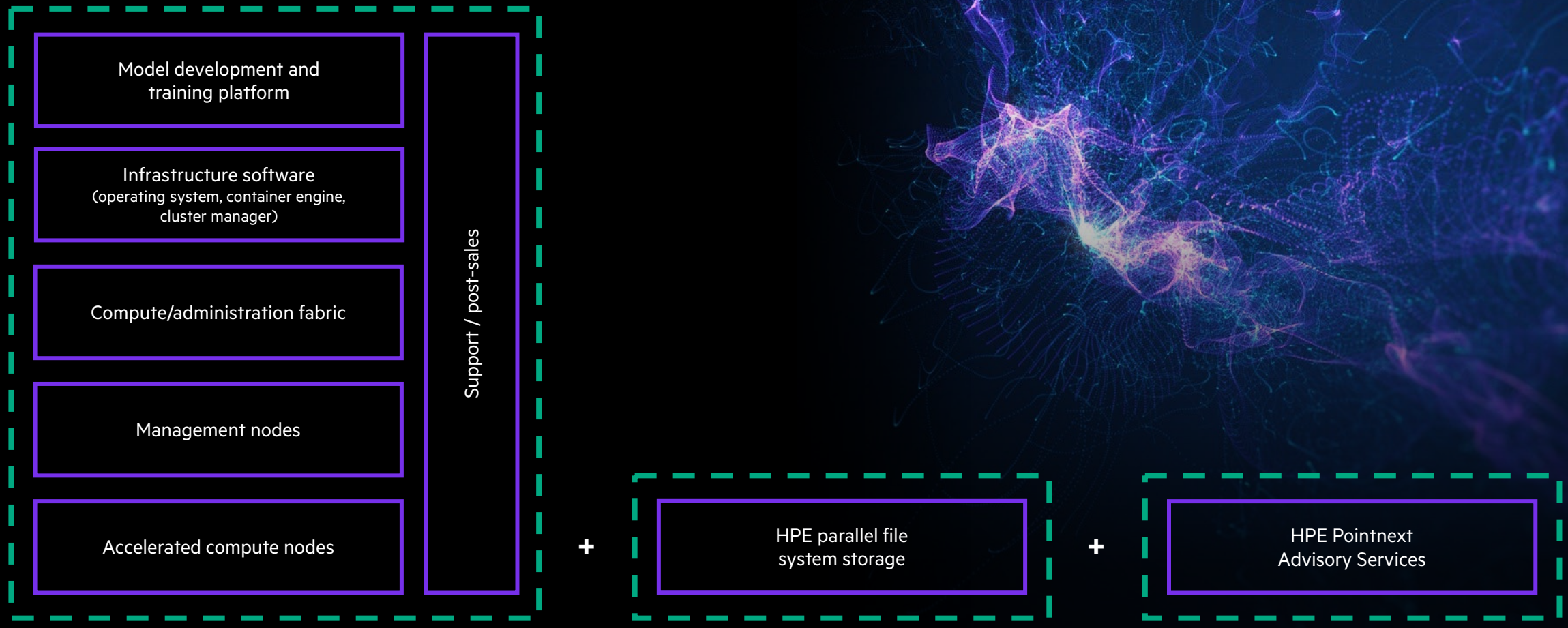
Seamlessly integrated components ⟶ Dramatically easier to use

[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



Core AI development system

Storage building block (optional)

Services building block (optional)

STANDARD HPE MACHINE LEARNING DEVELOPMENT SYSTEM OFFERING

Small system (4 nodes)

Infrastructure

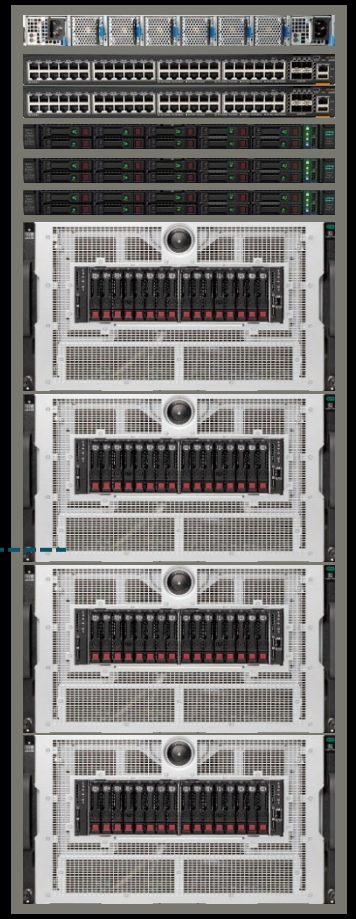
1x IB HDR switch

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)



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)

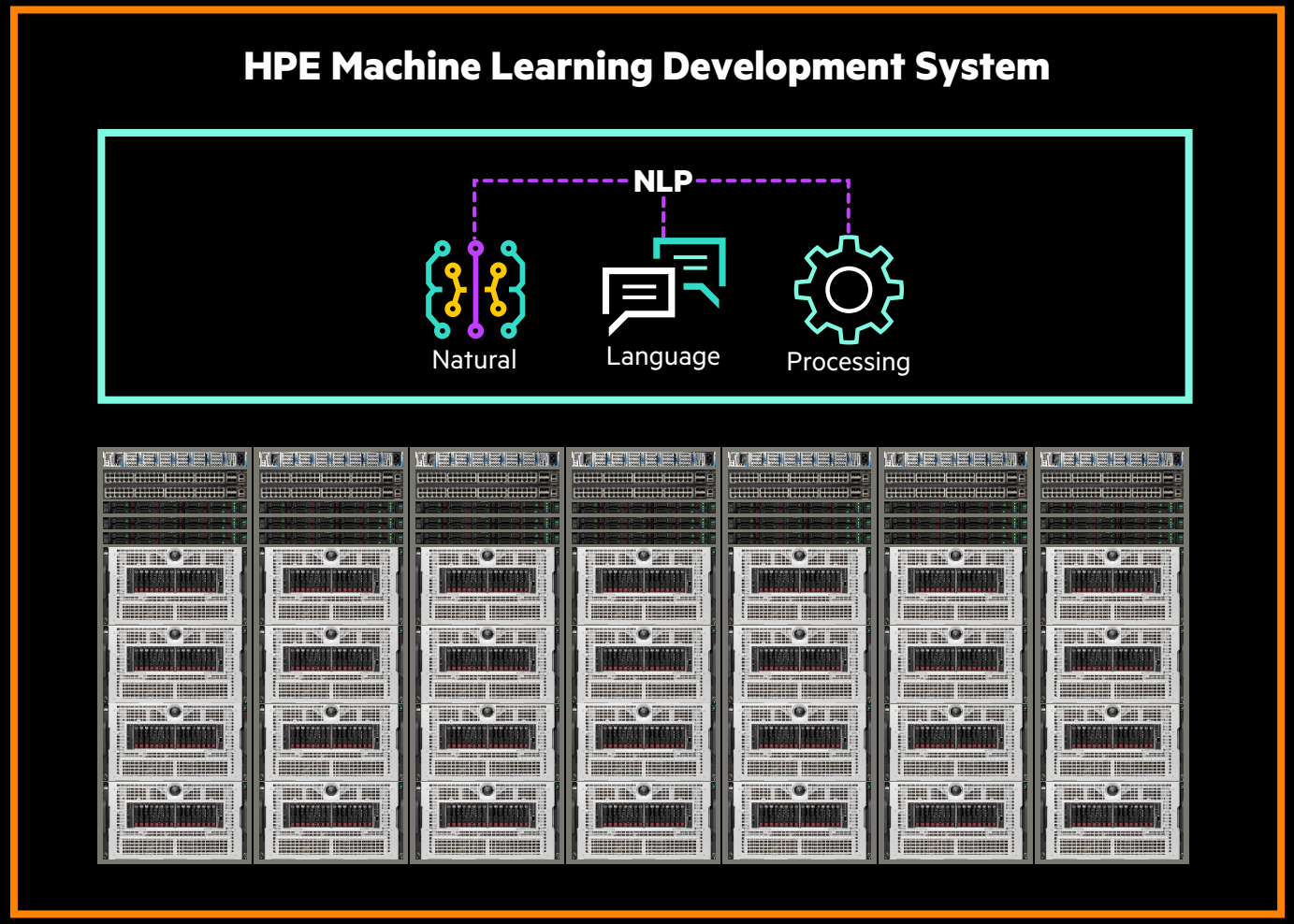
Services and software stack[1]

Services and support	Pre-installed, configured and benchmarked in factory	HPE Pointnext Services
	On-site system start up and validation (including software)	
	"Getting Started" one-day workshop	
	Full solution support	
Software	Training platform	HPE Machine Learning Development Environment
	Container runtime	Docker
	Cluster manager	HPE Performance Cluster Manager
	Operating system	Red Hat Enterprise Linux®

[1] Services design is WIP with HPE Pointnext Services

SOLUTION IN ACTION

HPE Machine Learning Development System delivers outcomes that matter



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