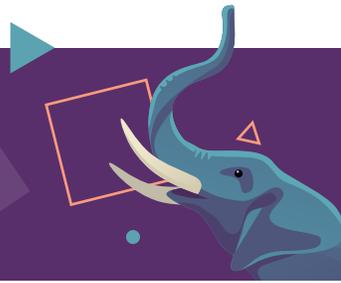# Pachyderm

Industrializing machine learning with NVIDIA DGX Systems and Pachyderm

Modern MLOps Challenges

When it comes to MLOps, no one tool can do it all. It takes an array of best-of-breed tools to automate and operationalize the entire ML lifecycle. Some of the workflow challenges faced by ML engineers and data scientists include managing:

- **Long lead times for production ML systems**
  - Production-ready ML applications take too long to implement and update, from creating infrastructure to writing code.

- **Tracking and processing data is necessary but expensive**
  - Rapidly increasing storage and processing costs hinder the ability to reliably track training data and intermediate results.

- **Data debt increases costs geometrically**
  - As old data is reprocessed with new data, processing costs increase geometrically over time. Companies are paying compound interest on this data debt.

That's why we're proud to be a member of the NVIDIA DGX-Ready Software program, to improve data science productivity, accelerate AI data pipeline and workflow, and improve accessibility and utilization of resources. Pachyderm is validated and certified with DGX systems, which takes the guesswork out of deploying this critical management layer for data science teams.

Pachyderm Overview

Pachyderm provides the data layer that allows machine learning teams to productionize and scale their machine learning lifecycle. With Pachyderm's industry leading data versioning, pipelines and lineage teams gain data driven automation, petabyte scalability and end-to-end reproducibility. Teams using Pachyderm get their ML projects to market faster, lower data processing and storage costs, and can more easily meet regulatory compliance requirements.
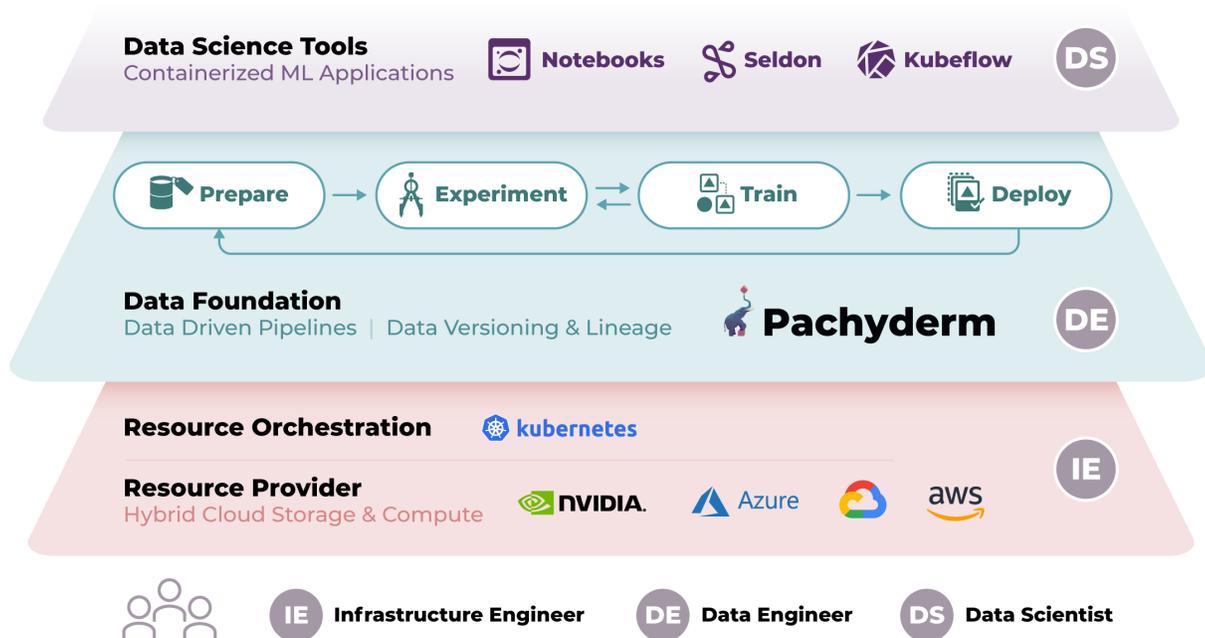
NVIDIA DGX Systems Overview

NVIDIA DGX™ A100 is the universal system for all AI workloads, offering unprecedented compute density, performance, and flexibility in the world's first 5 petaFLOPS AI system. NVIDIA DGX A100 features the world's most advanced accelerator, the NVIDIA A100 Tensor Core GPU, enabling enterprises to consolidate training, inference, and analytics into a unified, easy-to-deploy AI infrastructure that includes direct access to NVIDIA AI experts.

Benefits of Pachyderm and NVIDIA DGX Systems

Together, Pachyderm and NVIDIA DGX systems provide an integrated solution, purpose-built for the world's most complex enterprise ML/AI challenges.

- Faster time-to-solution through accelerated machine learning life cycles

- Get better model accuracy by being able to run larger datasets than ever before

- Add business value through increased quality of AI models and results

Pachyderm and NVIDIA DGX Systems Machine Learning Workflow

Data Science Tools
Containerized ML Applications   ⬚ Notebooks   ⧗ Seldon   ◆ Kubeflow   DS

Prepare → Experiment ⇄ Train → Deploy

Data Foundation
Data Driven Pipelines | Data Versioning & Lineage   🐘 **Pachyderm**   DE

Resource Orchestration   ⎈ kubernetes   IE

Resource Provider
Hybrid Cloud Storage & Compute   ▨ NVIDIA.   △ Azure   ☁   aws

👥   IE Infrastructure Engineer   DE Data Engineer   DS Data Scientist

About Pachyderm

Pachyderm is the data foundation for machine learning. Pachyderm provides industry leading data versioning, pipelines and lineage that allow data science teams to automate the machine learning lifecycle and optimize their machine learning operations (MLOps).

With investment from Benchmark, Microsoft M12, and others, Pachyderm, Inc. offers a user-deployed Pachyderm Enterprise Edition, a hosted SaaS Pachyderm Hub Edition and an open source Pachyderm Community Edition.

Pachyderm helps customers get their ML and AI projects to market faster, lower data processing and storage costs, and supports strict data governance requirements through data driven automation, petabyte scalability and end-to-end reproducibility.

Contact Pachyderm

To learn more about the Pachyderm machine learning solutions, contact us:

info@pachyderm.com · **888-338-9597** · **www.pachyderm.com**