The Fastest Path to Production AI
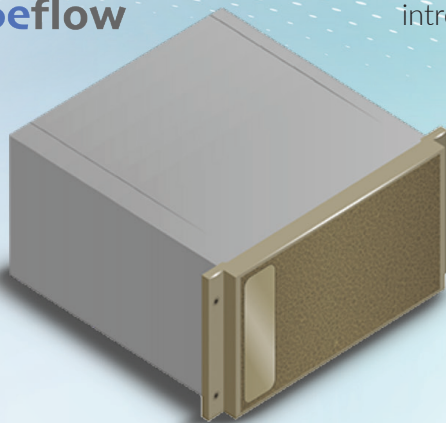
MLOps is Hard. The tools and technologies required for supporting AI initiatives are rapidly changing, and creating platforms that will evolve and scale is no easy feat. Supporting machine learning development often consists of deploying specialized infrastructure components that are different for data engineering, model training, and AI application serving. AI infrastructure is distinct from traditional IT infrastructure and providing resources to development teams in a way that does not create obstacles requires a new operational model. Leveraging public cloud offerings may provide some agility though there are always tradeoffs and controlling costs is not easy when resource demands are 24x7.

Kubeflow

ePlus Technology, a leading provider of AI optimized infrastructure, has partnered with Patterson Consulting, a leader in professional services for production machine learning, to simplify MLOps with the introduction of the **AI Accelerator Bundle**. This innovative solution built on the NVIDIA DGX A100 is combined with the Kubeflow machine learning platform to provide **AI as a Service** capabilities. A single system can be shared efficiently among a team of developers and as demands grow additional DGX systems can be added to easily scale.

Deployment and on boarding services are included so that you can go from **Zero to Operational** in a matter of days. To help you get results faster, consulting services are included that can be used for data engineering or data science work.

AI Accelerator Bundle
Components

+ (1) **NVIDIA DGX A100** 320GB or 640GB System
+ **Kubeflow** and **MicroK8s** Pre-Installed
+ Half-Day **Administrative Training** Session
+ 2-Day End-User Training Session
+ (40) Hours of **Data Engineering** and **Data Science** Consulting Services
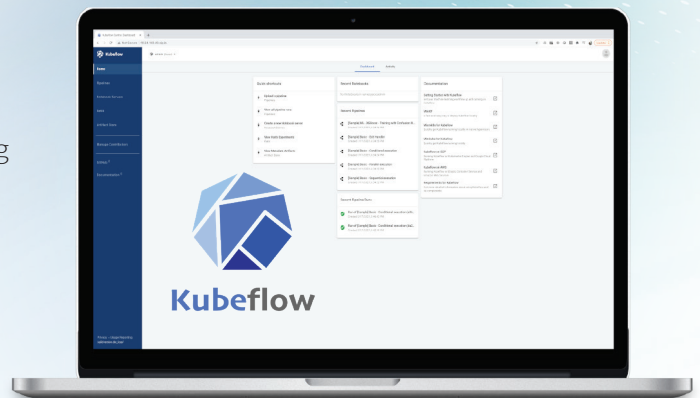+ (5) Copies of **O'Reilly's Kubeflow Operations Guide**

NVIDIA DGX A100:

+ 5 PFLOPS AI system, built on the NVIDIA A100 GPU
+ Up to 20X Higher GPU Performance vs Prior Generation Systems
+ One Universal Building Block for the AI Data Center
+ Up to (56) Independent Partitioned GPU Instances
+ Data Prep, Training, and Inference
+ Replace 100s of Servers with a Single DGX A100

Kubeflow:

+ Portable End-to-End ML Stack Orchestration
+ Self Service Multi-User, Multi-Project Support
+ Project Management with Detailed Monitoring
+ Pipelines with Reusable Components Across Projects
+ Popular ML Model Training Tools
+ Multi-Framework ML Serving

Kubeflow

For more information and to Get Started visit us at http://eplus.com/AIAC or email us at AI@eplus.com.

e+

Where Technology Means More®

13595 Dulles Technology Dr, Herndon, VA 20171-3413
888.482.1122 | tech@eplus.com | eplus.com | Nasdaq NGS: PLUS