Accelerating AI Workflows with MATLAB and NVIDIA NGC

Akhil Docca - Senior Product Marketing Manager at NVIDIA
Andy Thé – Partner Manager at MathWorks

MATLAB **EXPO** 2021

AGENDA

- NVIDIA and MathWorks Partnership

- What is the NVIDIA NGC catalog?

- How NVIDIA NGC helps with accelerating AI workflows?

- <u>Demo</u>: Deep Learning with MATLAB using NGC on AWS

- Conclusion and Call-to-Action

- Q&A

NVIDIA AND MATHWORKS COLLABORATION

INTEGRATE NVIDIA GPUS WITH MATLAB FROM EMBEDDED TO ENTERPRISE

GENERAL-PURPOSE TECHNICAL COMPUTING

NVIDIA GPU SUPPORT IN 800+ FUNCTIONS:

- MATLAB
- Deep Learning Toolbox
- Image Processing Toolbox
- Signal Processing Toolbox

No need for CUDA expertise!

ON-PREM AND CLOUD

MATLAB and NVIDIA GPU CLOUD

Instantly access on-premises and cloud GPUs with MATLAB and NGC

aws

Azure   Google Cloud

Alibaba Cloud   ORACLE CLOUD

EMBEDDED SYSTEMS

OPTIMIZED CUDA GENERATION FROM MATLAB CODE

MATLAB

GPU Coder

CUDA

cuDNN, cuSolver, cuBLAS TensorRT

AI HAS REACHED THE TIPPING POINT

Recommendation System

Conversational AI

Fraud Detection

Logistics

Medical Imaging

Autonomous Mobility
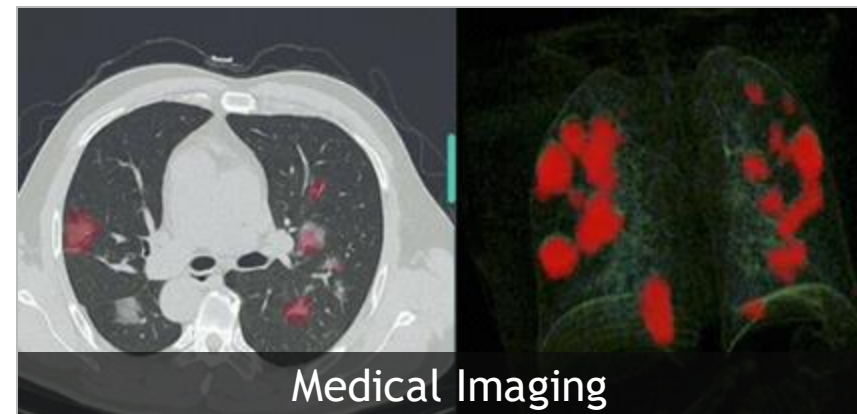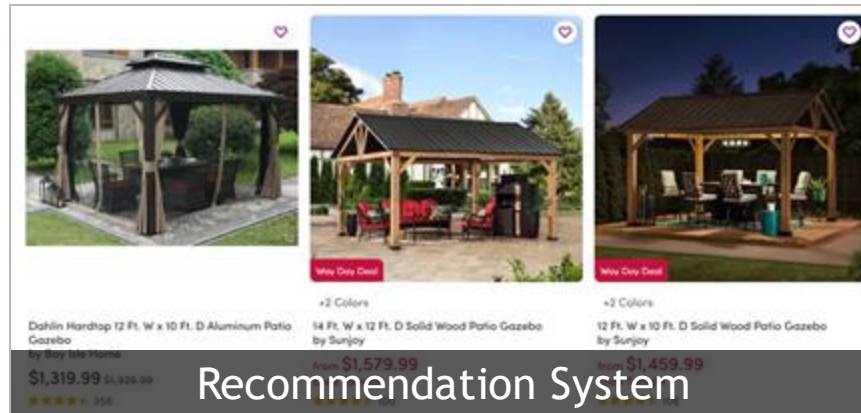
Over 50 Percent of Organizations Are Using AI in Production
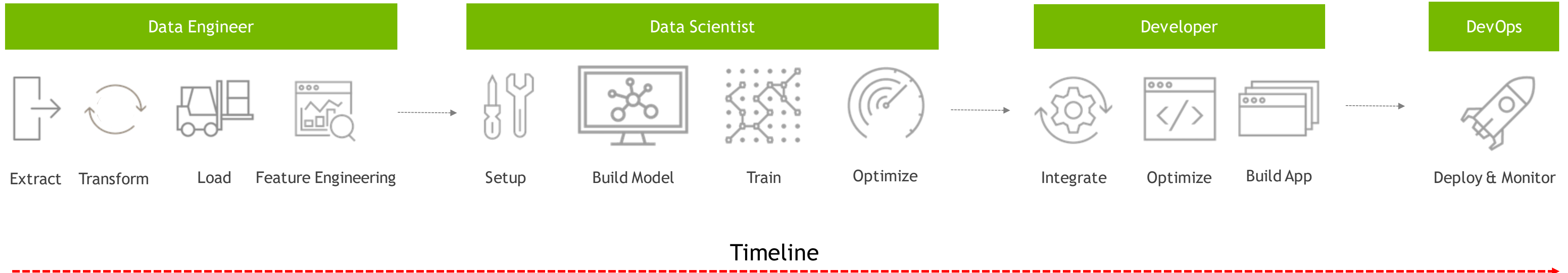
AI Adoption In The Enterprise 2020 O'Reilly Survey

NVIDIA.

POLL # 1

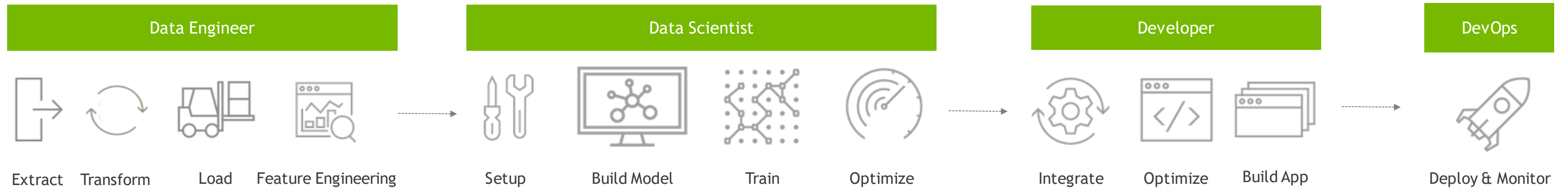Have you ever heard about NVIDIA NGC Catalog?
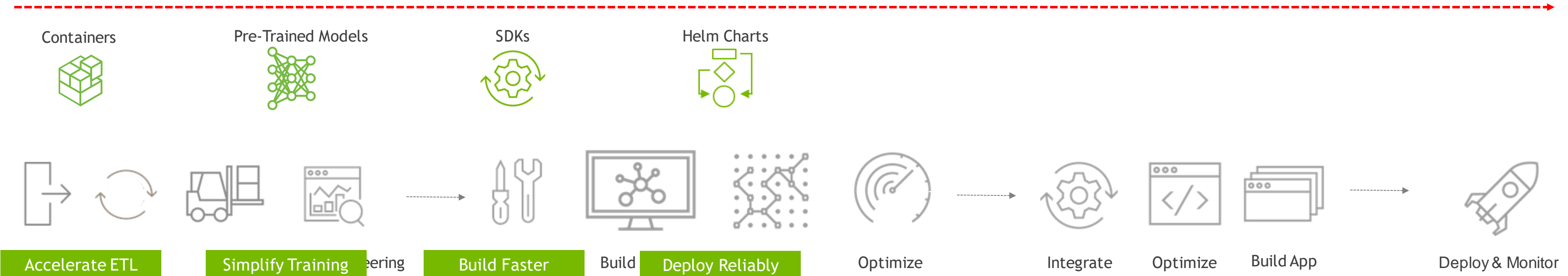
    a.   Yes

    b.   No

AI WORKFLOWS ARE COMPLEX

Data Engineer	Data Scientist	Developer	DevOps

Extract · Transform · Load · Feature Engineering → Setup · Build Model · Train · Optimize → Integrate · Optimize · Build App → Deploy & Monitor

Timeline

SIMPLIFYING AI WORKFLOWS WITH THE NGC CATALOG

Data Engineer	Data Scientist	Developer	DevOps

Extract   Transform   Load   Feature Engineering   Setup   Build Model   Train   Optimize   Integrate   Optimize   Build App   Deploy & Monitor

Timeline

Containers   Pre-Trained Models   SDKs   Helm Charts

Accelerate ETL   Simplify Training   eering   Build Faster   Build   Deploy Reliably   Optimize   Integrate   Optimize   Build App   Deploy & Monitor
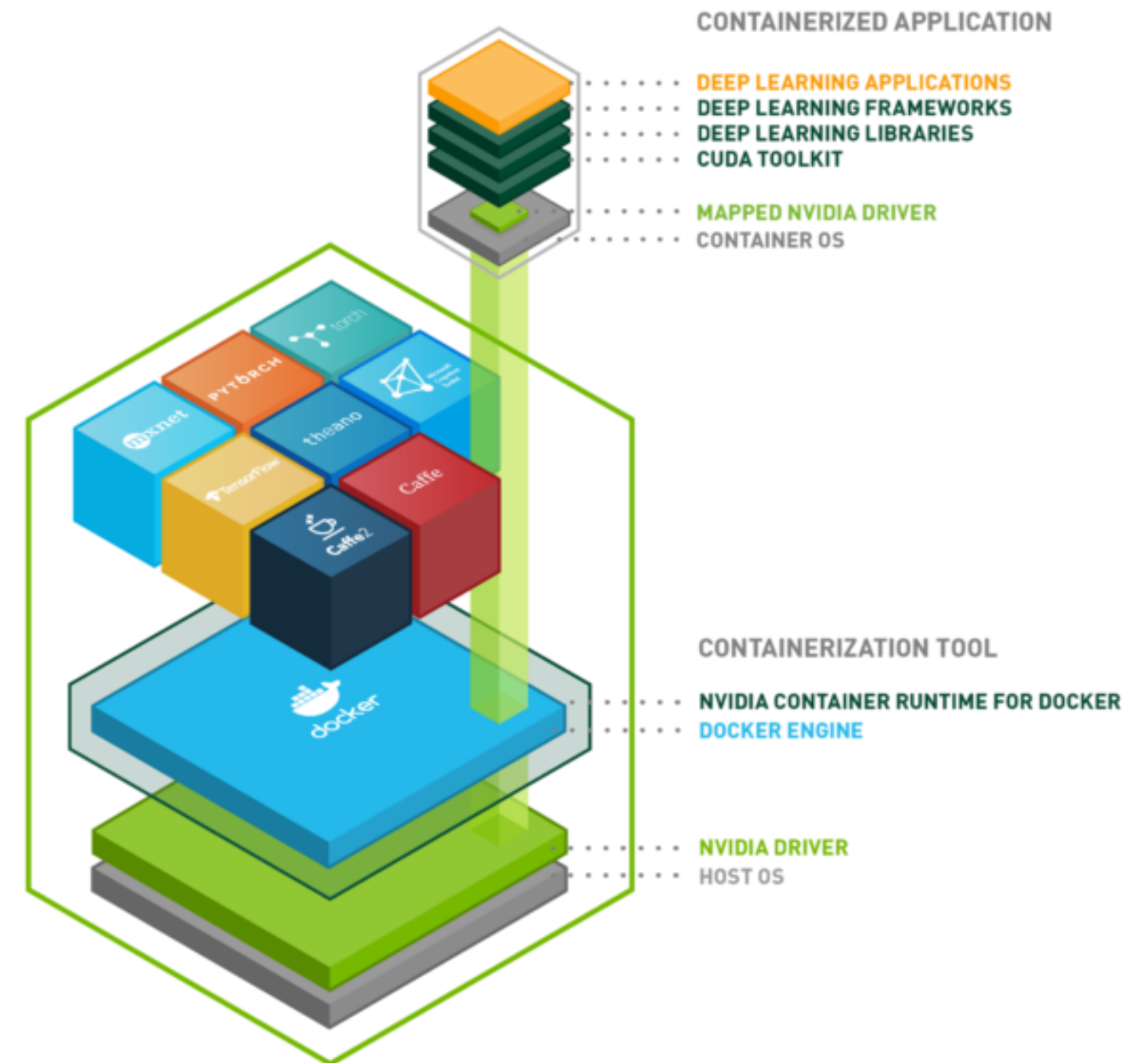
7

NVIDIA.

Poll # 2

Do you or your colleagues use containers today?
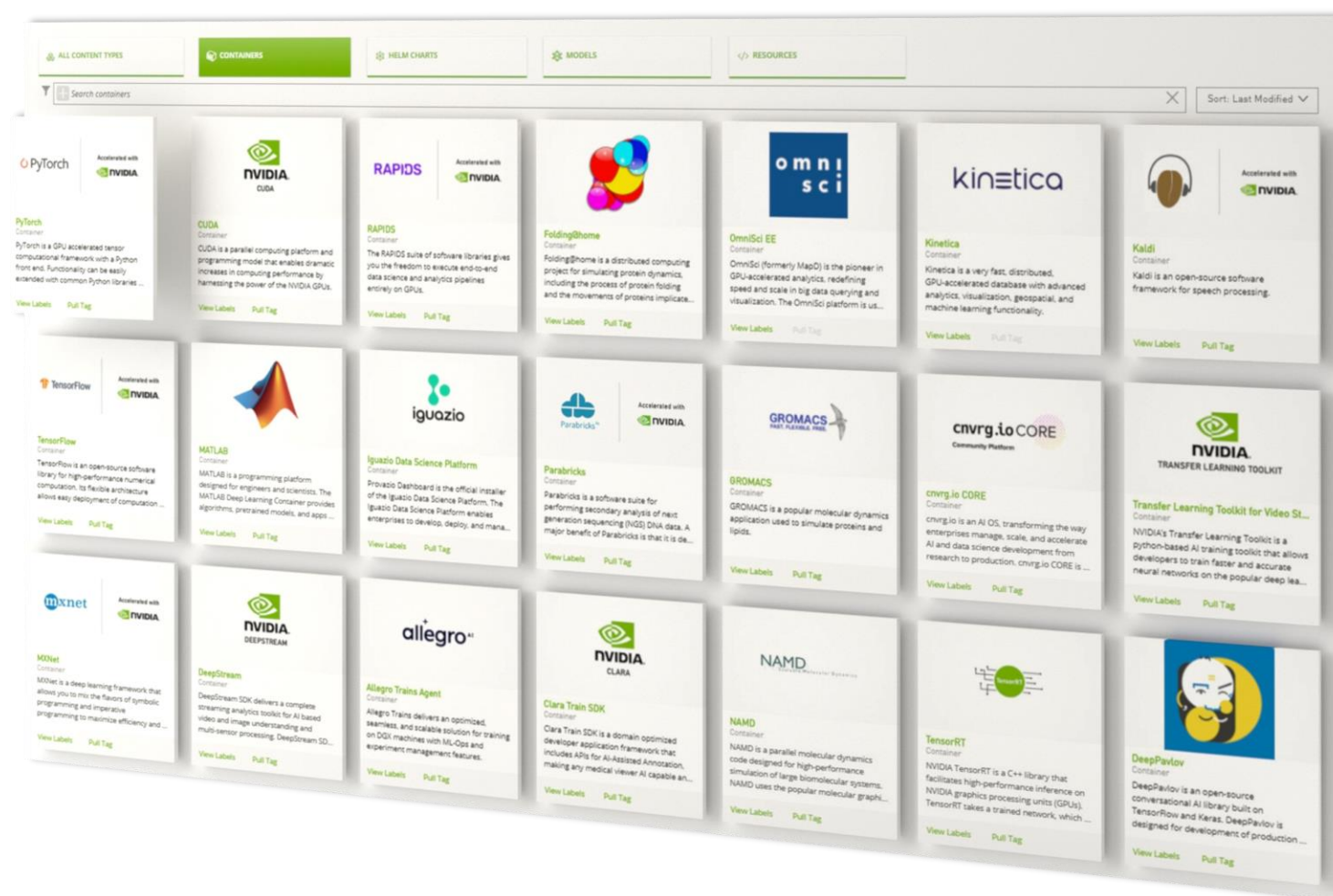
a. Yes

b. No

WHY CONTAINERS?

- Agnostic to the underlying OS

- Packages the application with its all dependencies

- Users can run multiple versions of the same application

- Easy to deploy, track and maintain

- Portability allows researchers to <u>collaborate</u> and <u>federate</u> scientific work

CONTAINERIZED APPLICATION

DEEP LEARNING APPLICATIONS
DEEP LEARNING FRAMEWORKS
DEEP LEARNING LIBRARIES
CUDA TOOLKIT

MAPPED NVIDIA DRIVER
CONTAINER OS

CONTAINERIZATION TOOL

NVIDIA CONTAINER RUNTIME FOR DOCKER
DOCKER ENGINE

NVIDIA DRIVER
HOST OS

NVIDIA.

NGC CONTAINERS ENABLE YOU TO FOCUS ON BUILDING AI

ENTERPRISE READY SOFTWARE

Scanned for CVEs, malware, crypto

Tested for reliability

Backed by Enterprise support

PERFORMANCE OPTIMIZED

Scalable

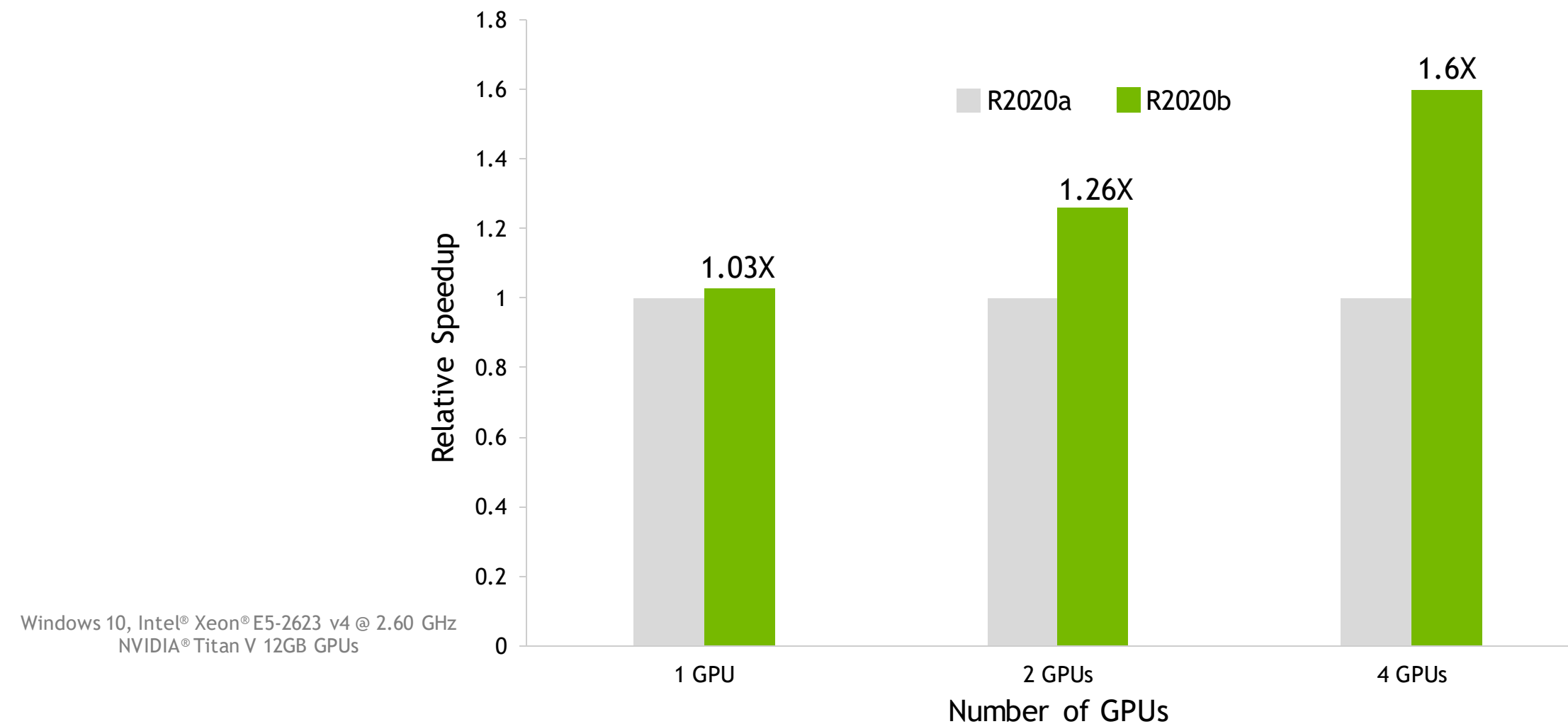Updated monthly

Better performance on the same system

DEPLOY ANYWHERE

Docker | cri-o | containerd | Singularity

Bare metal, VMs, Kubernetes

Multi-cloud, on-prem, hybrid, edge

DO WHAT YOU DO BEST, FASTER

R2020a ▢ R2020b ▢

Relative Speedup

- 1 GPU: 1.03X
- 2 GPUs: 1.26X
- 4 GPUs: 1.6X

Number of GPUs

Windows 10, Intel® Xeon® E5-2623 v4 @ 2.60 GHz
NVIDIA® Titan V 12GB GPUs

ENTERPRISE READY SOFTWARE

Scanned for CVEs, malware, crypto

Tested for reliability

Backed by Enterprise support

PERFORMANCE OPTIMIZED

Scalable

Updated Monthly

Better performance on the same system

DEPLOY ANYWHERE

Docker | cri-o | containerd | Singularity

Bare metal, VMs, Kubernetes

Multi-cloud, on-prem, hybrid, edge

FIND THE RIGHT FIT WITH CREDENTIALS

WIDE RANGE OF USE CASES

ResNet-50, SSD, MobileNet, VGG16

WaveGlow, BERT, NeMo

Wide & Deep, DLRM & many more

PRE-TRAINED MODELS

Faster training

Higher accuracy

Transparency through credentials

RESOURCES

Get started with code samples
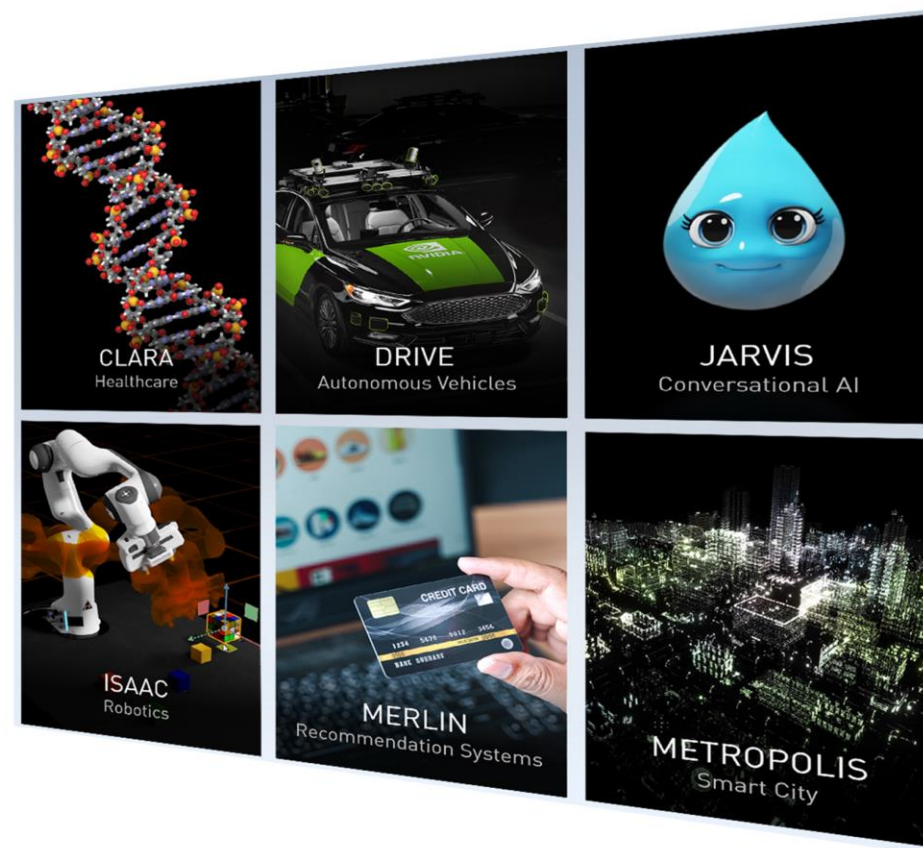
Customize NGC models

Reproduce with recipes

POLL # 3

How long have you been working with AI?

a. < 1 year

b. < 2 years

c. 2+ years

d. 5+ years

INDUSTRY APP FRAMEWORKS FOR END-TO-END AI WORKFLOWS

TRANSFER LEARNING TOOLKIT

Domain adaptability

Significantly reduce development time

TENSORRT

Optimizes for low latency and high-throughput

Integrated with major frameworks

TRITON

High performance inference

Supports multiple frameworks

Poll # 4

What AI application focus:
  a. NLP
  b. Computer vision
  c. Speech
  d. Recommender systems
  e. Other

NGC PRIVATE REGISTRY
Secure, Cloud-hosted Platform to Collaborate, Build, Deploy Faster

MANAGE ACCESS

Flexible role-based access control

SECURE SOFTWARE

Automatic container scanning

COLLABORATE

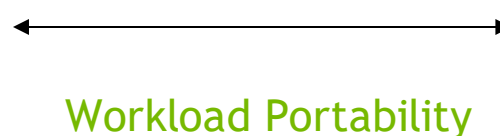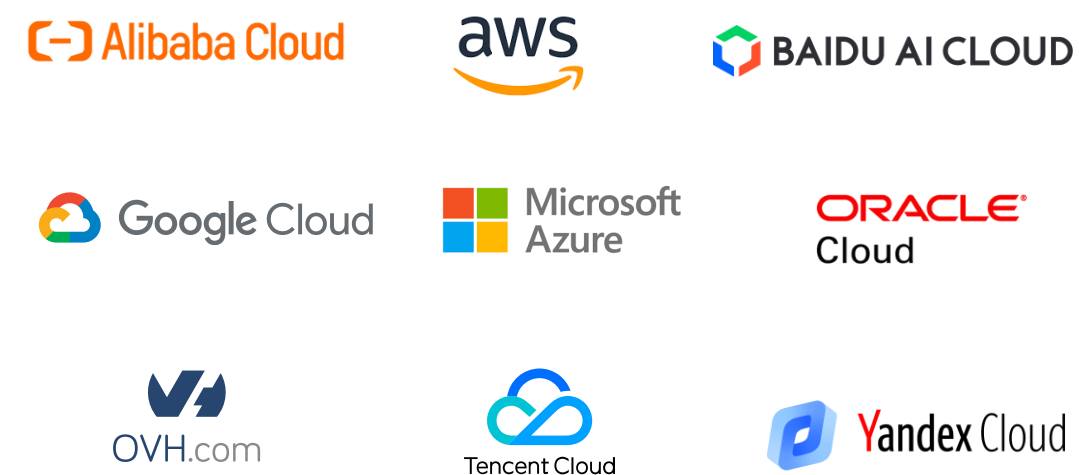Share containers, models and helm charts with your teams

nvidia.

DEPLOY ANYWHERE

Unified Experience Across On-Prem and Cloud

OEM Systems Certified by NVIDIA

Public Cloud Providers

Workload Portability

PERFORMANCE OPTIMIZED | TESTED FOR RELIABILITY | SCALABLE | SECURE | PORTABLE

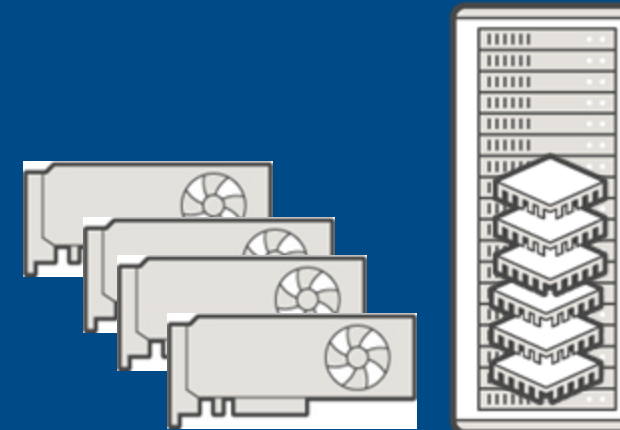# DEMO – DEEP LEARNING WITH MATLAB CONTAINER ON AWS

Data Preparation

Access and prepare large volumes of data

Design AI Network

Interactive tools for designing complex networks

AI Model Tuning

Accelerating training with on demand GPUs

Deployment

Deploy validated models anywhere

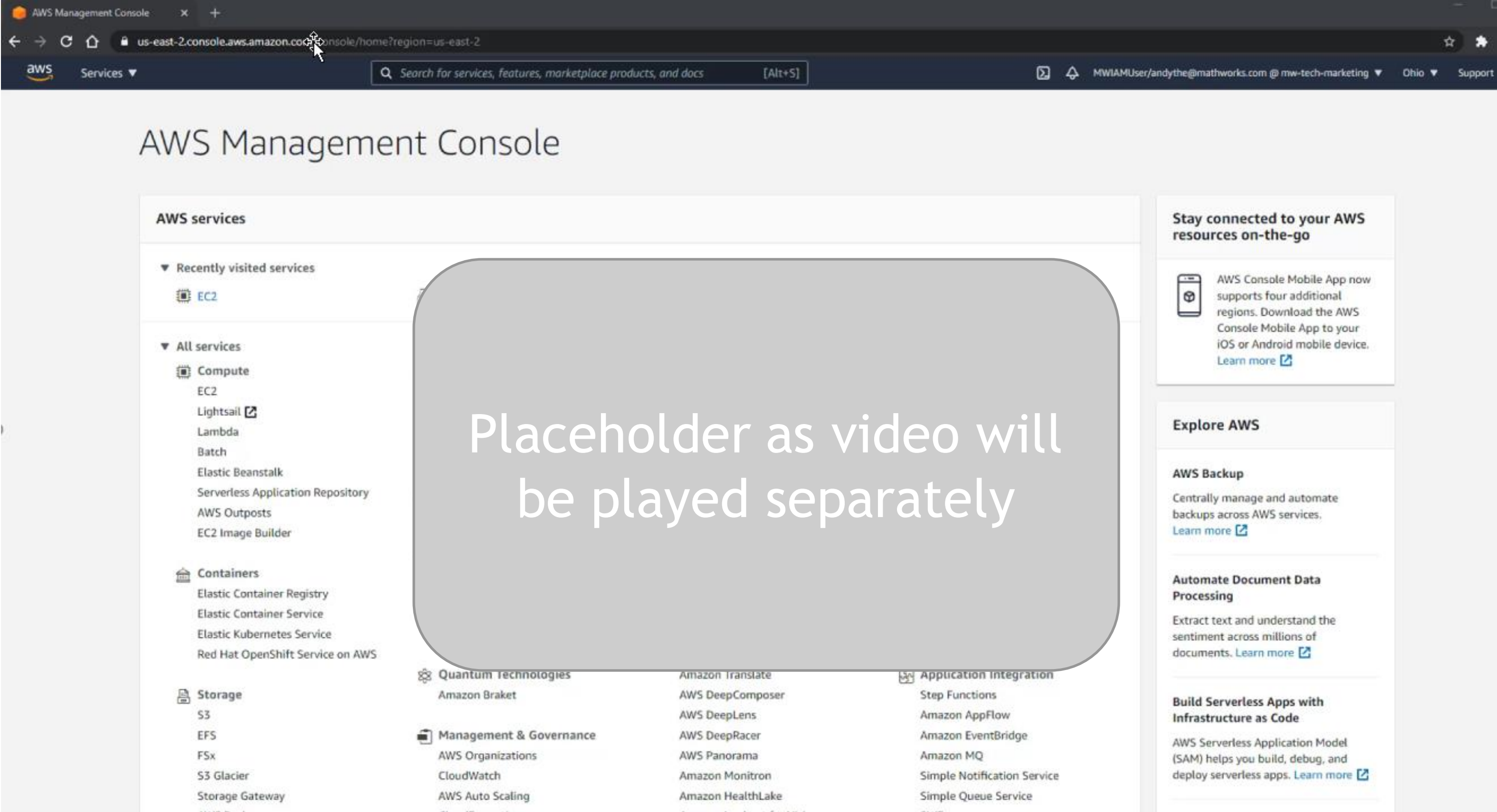# STEPS TO RUN MATLAB NGC DOCKER CONTAINER ON AWS

1. Create EC2 instance on AWS

2. Use PuTTY to connect to the EC2 instance

3. Pull & run the MATLAB container from the NGC catalog

4. Login into your MathWorks.com account

5. Run MATLAB deep learning examples

Check out the blog for a detailed walkthrough:
- tinyurl.com/3urvtyp8 or
- https://blogs.mathworks.com/deep-learning/2021/05/03/ai-with-matlab-ngc

NVIDIA.

Demo Video: MATLAB+NGC+AWS

Placeholder as video will be played separately

ACCELERATING TRAINING WITH MULTIPLE GPUS ON IMAGENET

- 150 GB

- >1.2M images

- 50k test images

- 1,000 categories

- ~3.5x faster / Epoch
  - 4 vs 1 GPU - V100

ResNet-50 training on ImageNet data-set
Training time per Epoch (V100)

~3.5x Faster w/ 4 GPUs

4 GPUs	45
2 GPUs	85
1 GPU	162

0   20   40   60   80   100   120   140   160   180

■ Minutes
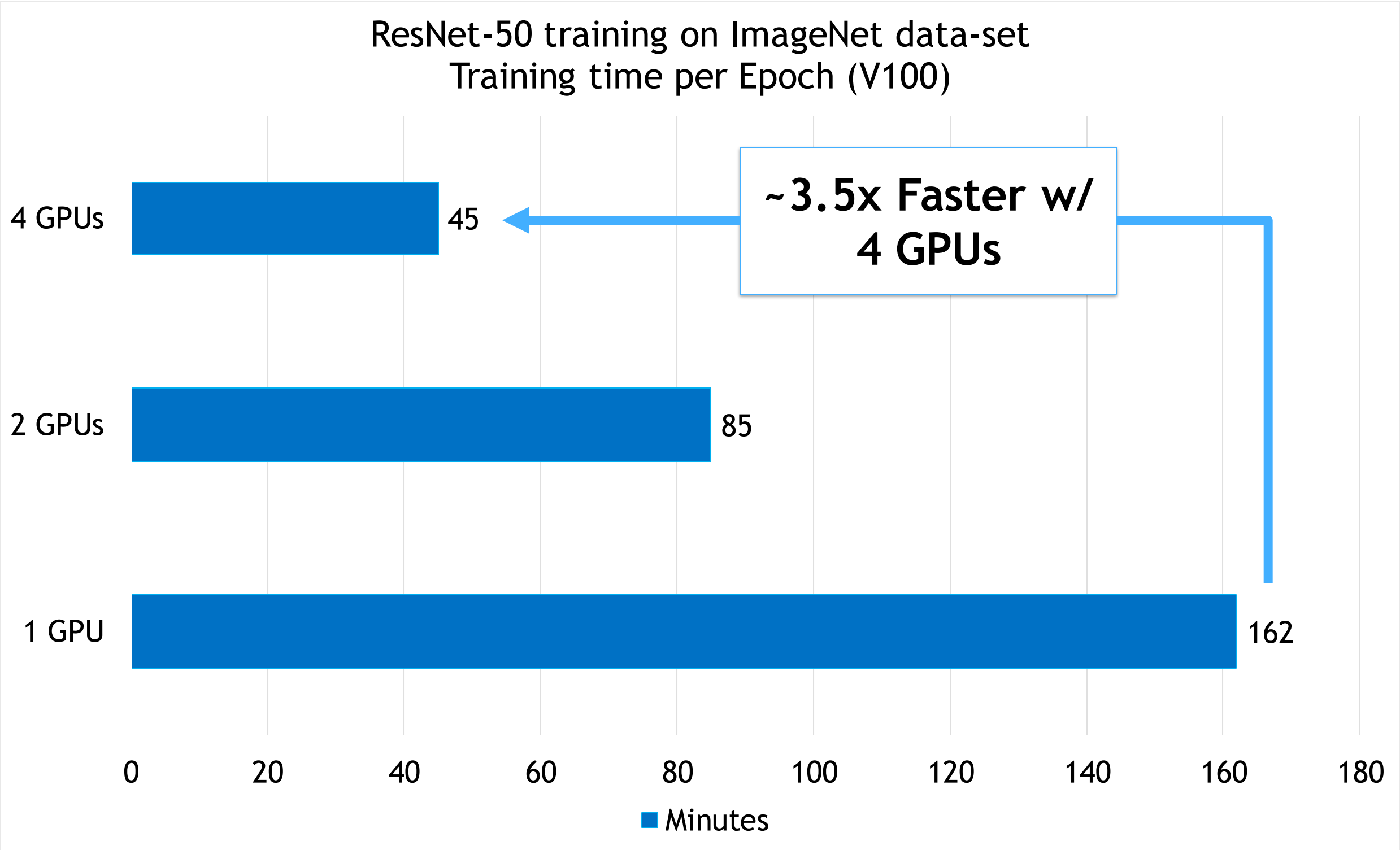
 NVIDIA.

START BUILDING AI FASTER WITH THE NGC CATALOG

Get started today with MATLAB and NGC:
https://ngc.nvidia.com/catalog/containers/partners:matlab

Check out the blog for a detailed walkthrough:

http://tinyurl.com/3urvtyp8

Learn more at nvidia.com/ngc