
NVIDIA
ACCELERATING AI AND HPC WITH
NGC AND MICROSOFT AZURE

AGENDA

Challenges with Complex Software

NGC and Microsoft Azure

Inside NGC Containers

About Azure

Demo

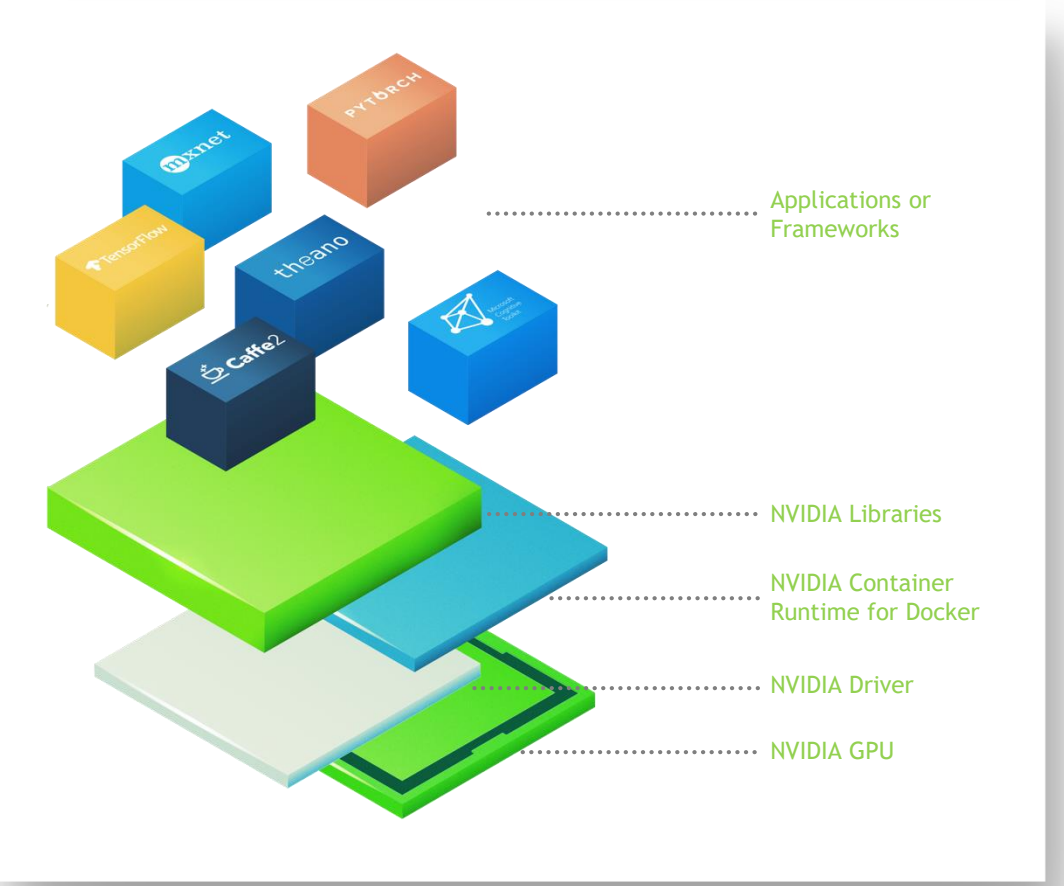
Q & A

CHALLENGES WITH COMPLEX SOFTWARE

Current DIY GPU-accelerated AI and HPC deployments are **complex** and **time consuming** to build, test and maintain

Development of software by the community is moving **very fast**

Requires high level of **expertise** to manage driver, library, framework dependencies



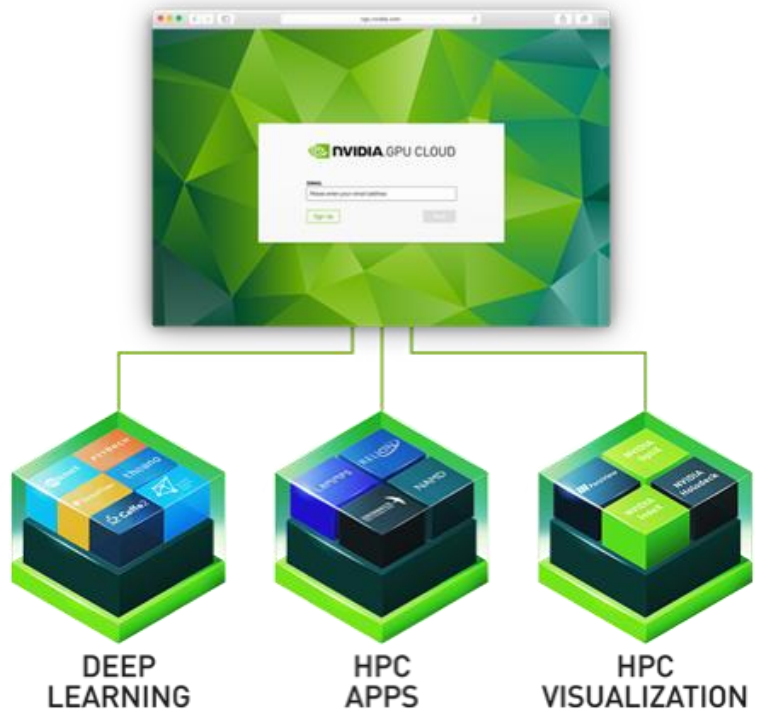
NVIDIA GPU CLOUD (NGC)

Simple Access to GPU-Accelerated Software

Discover 49 GPU-Accelerated Containers
Deep learning, HPC applications, HPC visualization tools, and partner applications

Innovate in Minutes, Not Weeks
Optimized, pre-configured, and ready-to-run

Run on a Variety of NVIDIA GPUs
Scale to the right size for your project



GPU-ACCELERATED CONTAINERS

Get Up and Running Immediately

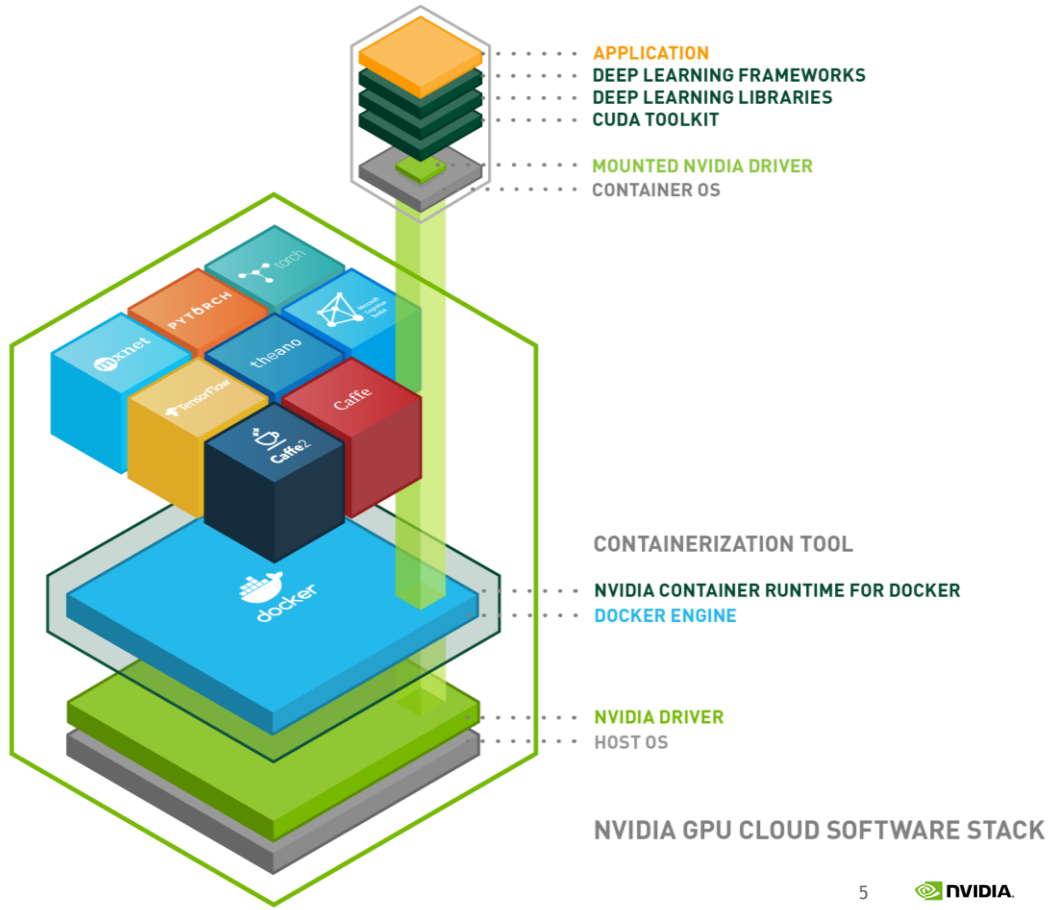
Tuned and **tested** to maximize performance

Cross-stack **optimizations**

Pre-integrated and ready-to-run

Frameworks and applications are **isolated**

Take advantage of the latest **NVIDIA GPU** instance types on Azure



CONTINUAL EXPANSION

Deep Learning	Machine Learning	HPC	HPC Visualization	NVIDIA/K8s	Partners
caffe	rapids	bigdft	index	Kubernetes on NVIDIA GPUs	chainer
caffe2		candle	paraview-holodeck		h20ai-driverless
cntk		chroma	paraview-index		kinetica
cuda		gamess	paraview-optix		mapd
Digits		gromacs			matlab
inferenceserver		lammps			paddlepaddle
mxnet		lattice-microbes			
pytorch		milc			
tensorflow		namd			
tensorrt		pgi-compilers			
tensorrtserver		picongpu			
theano		relion			
torch		vmd			

10 containers

October 2017

NEW CONTAINERS

49 containers

October 2018

USING NGC CONTAINERS

Benefits for a Wide Variety of Users

Data Scientists and Researchers

Eliminate setup time, focus on science and research

Developers

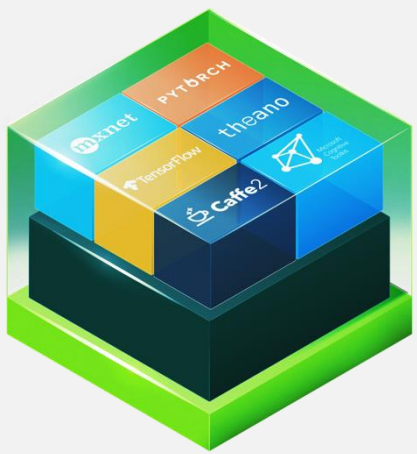
Work with the latest software with a known good starting point

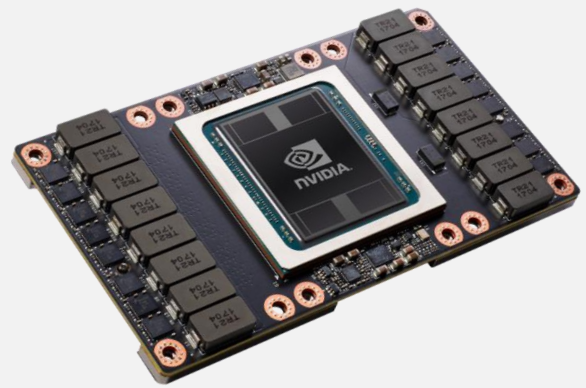
Sysadmins

Deploy to production immediately

NGC CONTAINERS ON MICROSOFT AZURE

Ready-To-Run Containers for NVIDIA GPU Instances on Azure





A GPU-enabled VM for every Scenario and Price

NC v2

	NC6s_v2	NC12s_v2	NC24s_v2	NC24rs_v2
Cores (Broadwell 2.6Ghz)	6	12	24	24
GPU	1 x P100	2 x P100	4 x P100	4 x P100
Memory	112 GB	224 GB	448 GB	448 GB
Local Disk	~700 GB SSD	~1.4 TB SSD	~3 TB SSD	~3 TB SSD
Network	Azure Network	Azure Network	Azure Network	Azure Network + InfiniBand

NC v3

	NC6s_v3	NC12s_v3	NC24s_v3	NC24rs_v3
Cores (Broadwell 2.6Ghz)	6	12	24	24
GPU	1 x V100	2 x V100	4 x V100	4 x V100
Memory	112 GB	224 GB	448 GB	448 GB
Local Disk	~700 GB SSD	~1.4 TB SSD	~3 TB SSD	~3 TB SSD
Network	Azure Network	Azure Network	Azure Network	Azure Network + InfiniBand

ND

	ND6s	ND12s	ND24s	ND24rs
Cores (Broadwell 2.6Ghz)	6	12	24	24
GPU	1 x P40	2 x P40	4 x P40	4 x P40
Memory	112 GB	224 GB	448 GB	448 GB
Local Disk	~700 GB SSD	~1.4 TB SSD	~3 TB SSD	~3 TB SSD
Network	Azure Network	Azure Network	Azure Network	Azure Network + InfiniBand

Introducing ND_v2: Next-Gen Volta GPU Compute

- Volta SXM GPU instances - NVIDIA Tesla V100 GPUs
- 8X NVIDIA V100 GPUs interconnected with NVLink mesh
- Excellent for accelerating machine training jobs and HPC
- Skylake based processor with premium storage support (SSD backed)
- Availability: Q4 CY2018
- Specs:
  - GPU Memory 16 GB
  - 300 GB/s GPU interconnect through NVLink

	ND40s_v3
Cores	40 cores
GPU	8 x V100 SXM
Memory	672 GB
Local Disk	~1.3 TB SSD
Network	Azure Network + NVLink GPU interconnect


DEMO: NGC CONTAINERS ON AZURE

ACCELERATE AI AND HPC WITH NGC AND MICROSOFT AZURE

Comprehensive Library of **GPU-Accelerated** Containers	**Ready-To-Run** on Microsoft Azure	Get Started in **Minutes**

Learn More:

nvidia.com/ngc **azure.microsoft.com**


Q & A


NVIDIA