

USING NVIDIA GPU CLOUD CONTAINERS ON NIMBIX CLOUD



GPU-Accelerated Innovation

Greg Crider
Senior Product Manager
NVIDIA



NVIDIA + NIMBIX PARTNERSHIP

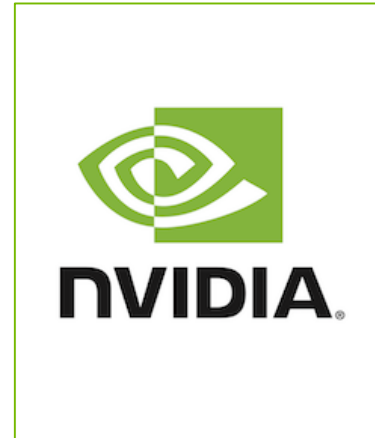
Working Together to Deliver Accelerated Computing

Nimbix was the first cloud provider to partner with NVIDIA

Has offered every NVIDIA compute GPU, including latest Volta V100 GPU

Only public cloud provider in the Americas to offer NVIDIA DGX-1 systems

NIMBIX

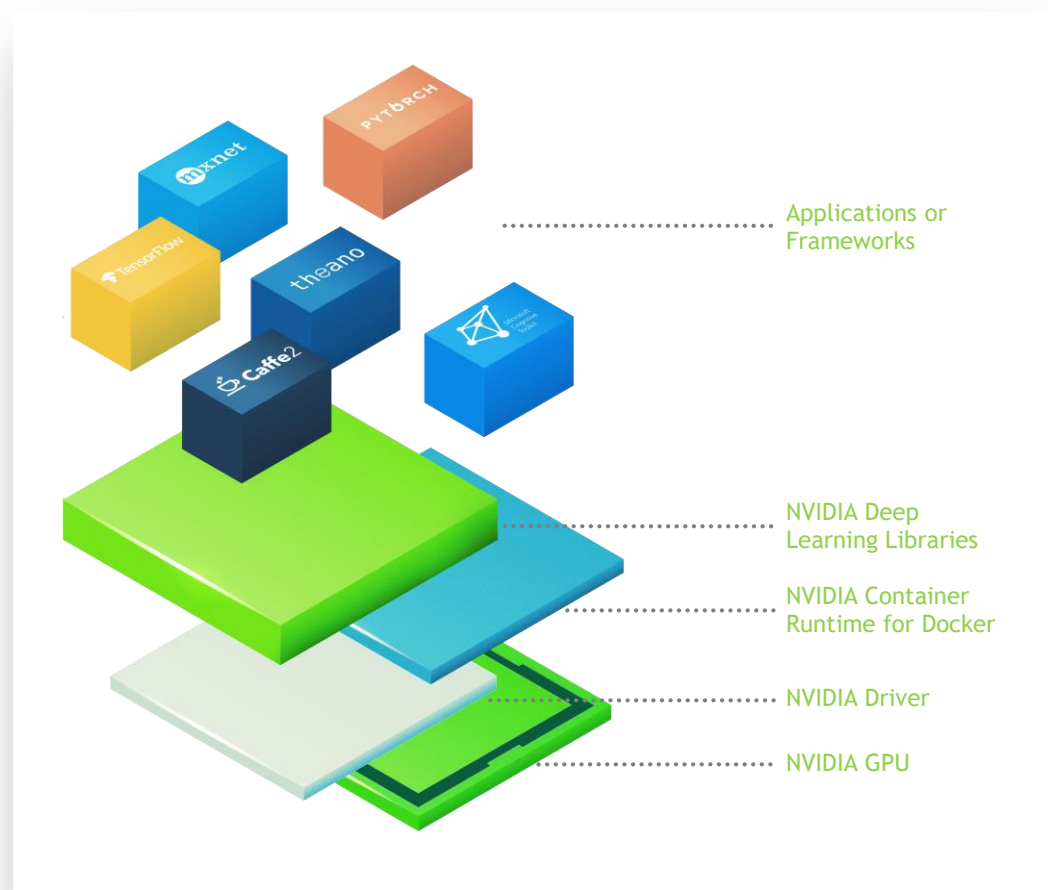


CHALLENGES WITH COMPLEX SOFTWARE

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

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

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

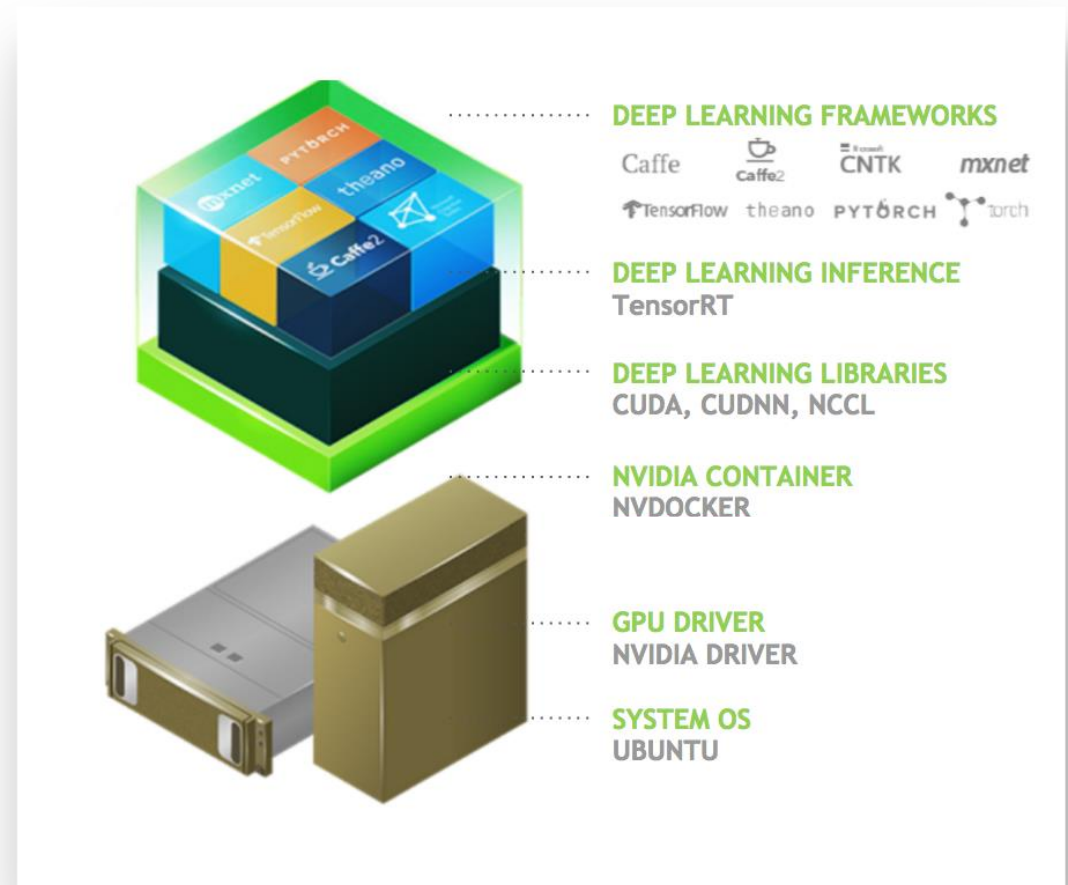


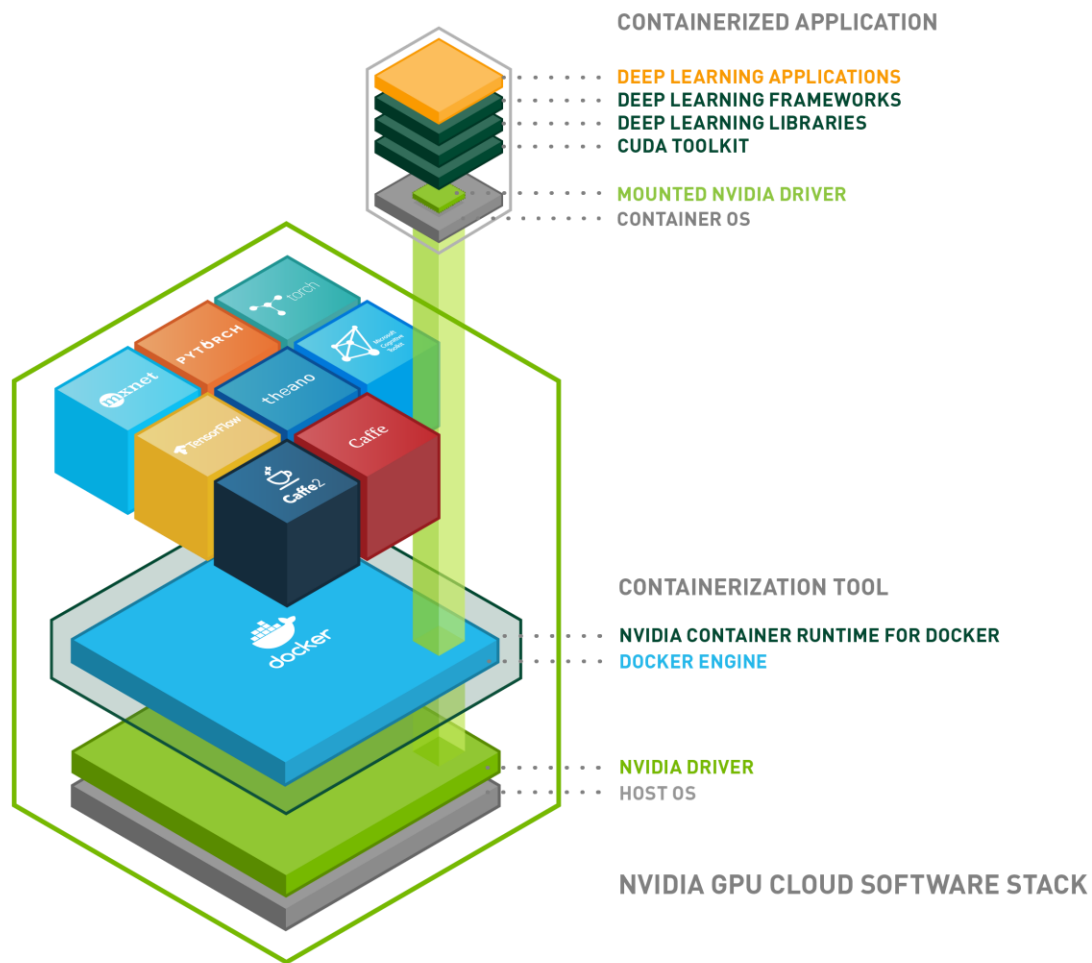
NGC DELIVERS A GPU OPTIMIZED ENVIRONMENT

Integrates the deep learning SDK, libraries, drivers and OS into a single environment

Removes all the **DIY complexity** of managing dependencies, updates

Stack runs in a VM **locally or in the cloud**





WHY CONTAINERS?

Benefits of Containers:

Simplify deployment of GPU-accelerated software, eliminating time-consuming software integration work

Isolate individual deep learning frameworks and applications

Share, collaborate, and test applications across different environments

NVIDIA GPU CLOUD

Simple access to GPU-accelerated software

Discover 30 GPU-Accelerated Containers

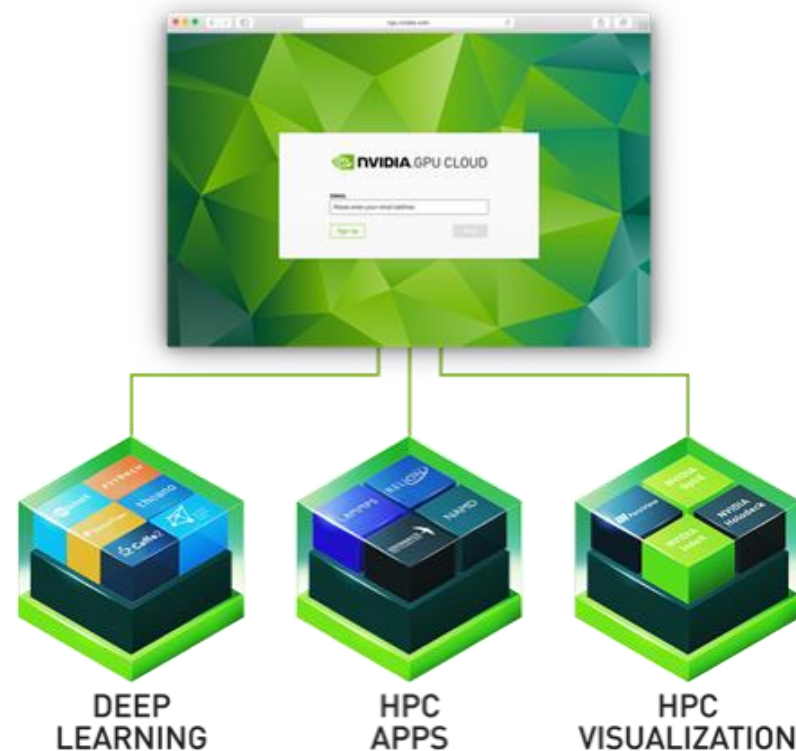
Deep learning, HPC, and partner applications

Innovate in Minutes, Not Weeks

Get up and running quickly

Access from Anywhere

Supports NVIDIA Volta or Pascal™ architecture GPUs on NVIDIA DGX Systems and supported cloud providers



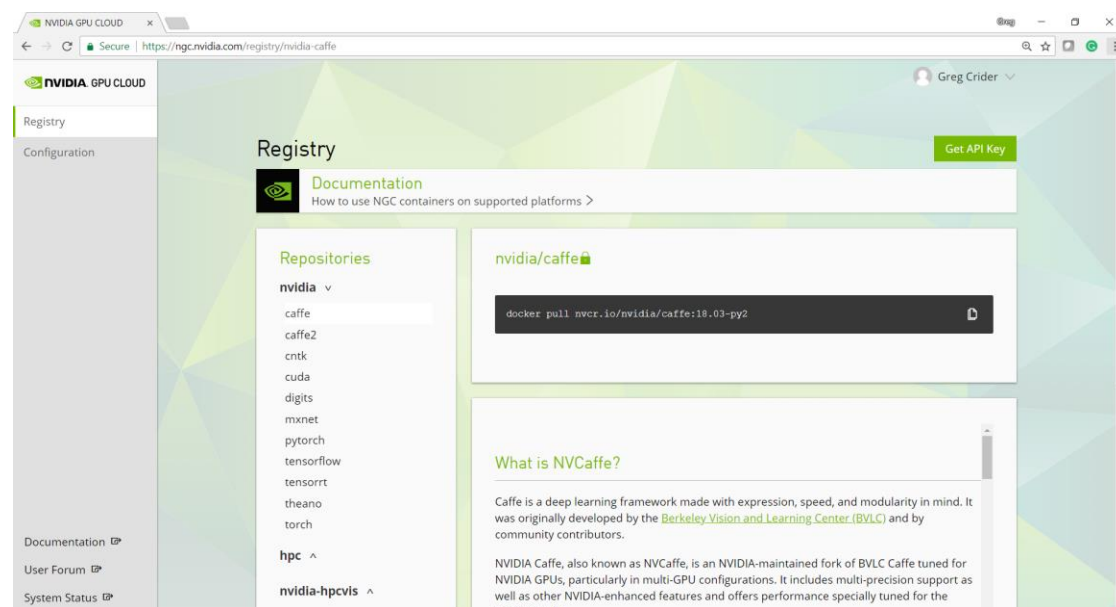
NGC CONTAINER REGISTRY

One stop shop for GPU optimized containers

Deep Learning Frameworks, drivers, OS, all pre-configured, tested, and optimized, updated monthly

Curated HPC and third-party applications

Updated monthly to deliver the highest performance frameworks in the industry



CURRENT NGC CONTAINERS

DEEP LEARNING

CAFFE
CAFFE2
CNTK
CUDA
DIGITS
MXNET
PYTORCH
TENSORFLOW
TENSORRT
THEANO
TORCH

HPC VIZ

PARAVIEW-HOLODECK
PARAVIEW-INDEX
PARAVIEW-OPTIX
INDEX
VMD

HPC

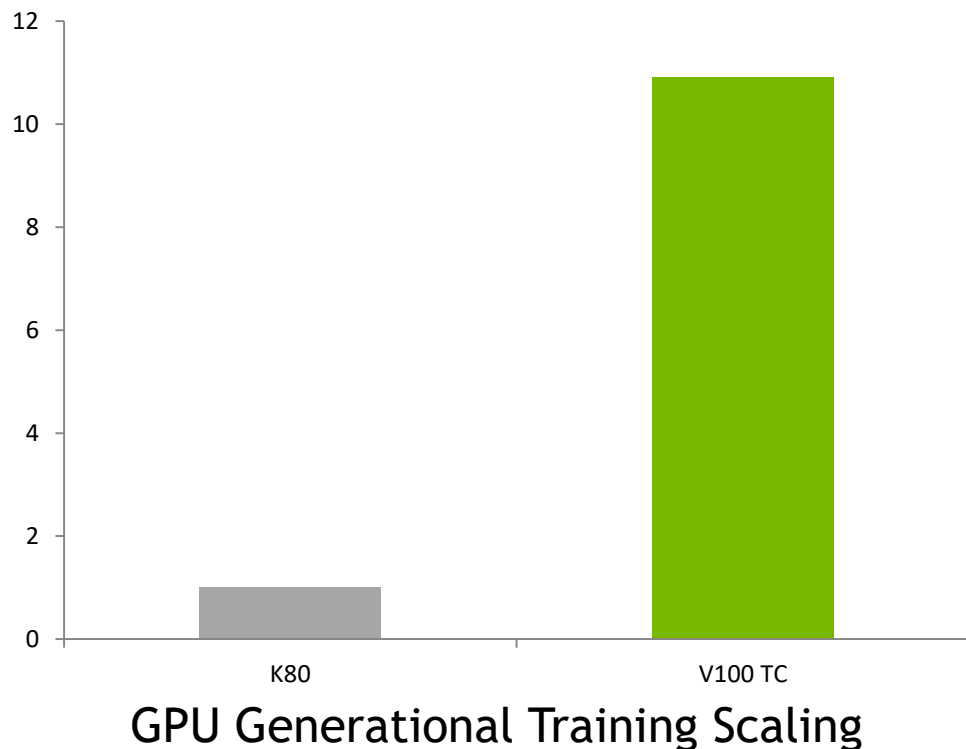
GAMESS
GROMACS
LAMMPS
NAMD
CHROMA
MILC
RELION
CANDLE
LATTICE MICROBES

PARTNER

H2O
MAPD
CHAINER
PADDLEPADDLE
KINETICA

MAXIMIZING PERFORMANCE ON VOLTA

Greater than 10x Performance on Volta vs K80



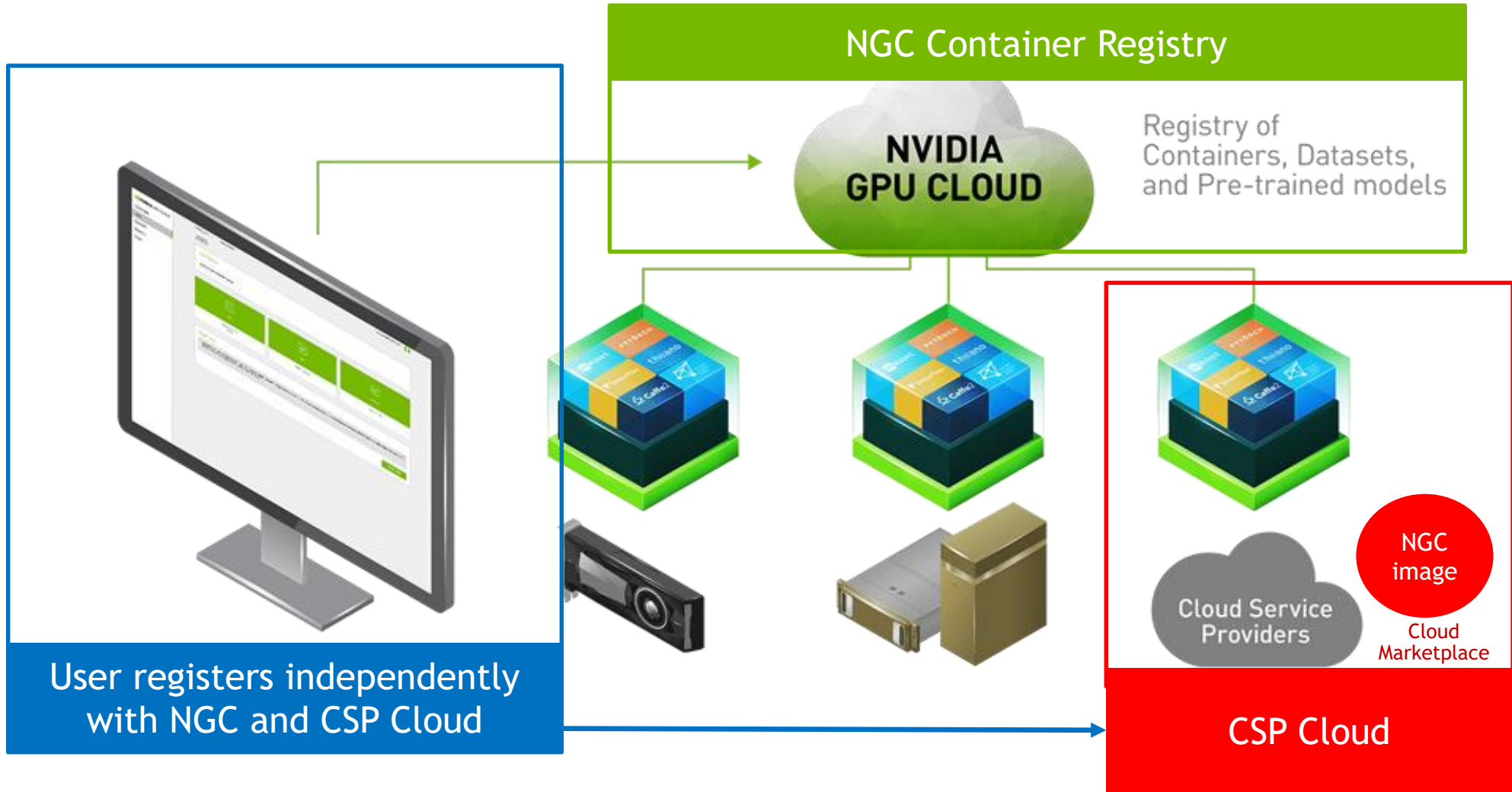
Frameworks on NGC have the latest NVIDIA optimizations for Volta

Cloud images from NVIDIA are tuned for maximum performance

Volta is >10x faster than K80

ResNet-152 Training, 8x K80 (16 GPUs total) compared with 8x V100 NVLink GPUs using NVIDIA 17.10 containers

NGC CONTAINERS RUNNING ON A CSP CLOUD



3 STEPS TO USING NGC ON A CSP CLOUD

CREATE A NGC ACCOUNT

User creates a free NGC account on NVIDIA's web site running in the US

www.nvidia.com/ngcsignup

CREATE A CSP CLOUD ACCOUNT AND RUN A GPU INSTANCE

User creates a CSP Cloud account and starts a GPU instance using the NGC image

DOWNLOAD NGC CONTAINERS

User running a GPU instance in the CSP Cloud downloads no-cost containers from the NGC Container Registry

The screenshot shows the 'NVIDIA GPU CLOUD' sign-up page. It has a green header with the NVIDIA logo and the text 'DEEP LEARNING EVERYWHERE, FOR EVERYONE'. Below the header, there's a 'Sign Up' section with a form. The form has fields for 'NAME' (John Smith), 'COMPANY' (Acme), 'EMAIL' (john@acme.com), 'ROLE' (Engineer), 'INDUSTRY' (Supercomputing), and 'COUNTRY' (United States). There are checkboxes for 'I agree to the NVIDIA ACCOUNT TERMS OF USE' and 'Send me NVIDIA GPU Cloud updates and news'. At the bottom, there are 'Cancel' and 'Sign Up' buttons.

The screenshot shows the 'NVIDIA Volta Deep Learning AMI' page. It has a green header with the NVIDIA logo and the text 'Sold by: NVIDIA Latest Version: 18.03.0'. Below the header, there's a 'Product Overview' section. The text describes the NVIDIA Volta Deep Learning AMI as an optimized environment for running deep learning frameworks. It mentions that the Docker containers available on the NGC container registry are tuned, tested, and certified by NVIDIA. It also mentions that the AMI is optimized for AWS PS1 Volta instances. There's a 'Highlights' section with bullet points: 'Provides AI researchers with fast and easy access to NVIDIA Volta GPUs in the cloud, with performance-engineered deep learning framework containers that are fully integrated, optimized, and certified by NVIDIA.', 'With 640 Tensor Cores, NVIDIA Volta is the world's first GPU to break the 100 teraflops (TFLOPS) barrier of deep learning performance.', and 'NVIDIA accelerates innovation by eliminating the complex do-it-yourself task of building and optimizing a complete deep learning software stack tuned specifically for GPUs.' There's also a 'Pricing' section showing a 'Typical Total Price' of '\$3.06/hr'.

The screenshot shows a terminal window with a command to pull a Docker container from the NGC Container Registry. The command is: `docker pull nvidia/tensorflow:17.10`. The output shows the progress of the pull, including the extraction of the image and the download of the container. The output also shows the size of the container (12.7MB/47.2MB) and the size of the image (66.39MB/427.6MB).

NVIDIA GPU CLOUD

GPU-accelerated containers for deep learning, HPC, and HPC visualization



**GPU-Accelerated
Containers**



**Innovate In Minutes,
Not Weeks**



**Access from
Anywhere**



