

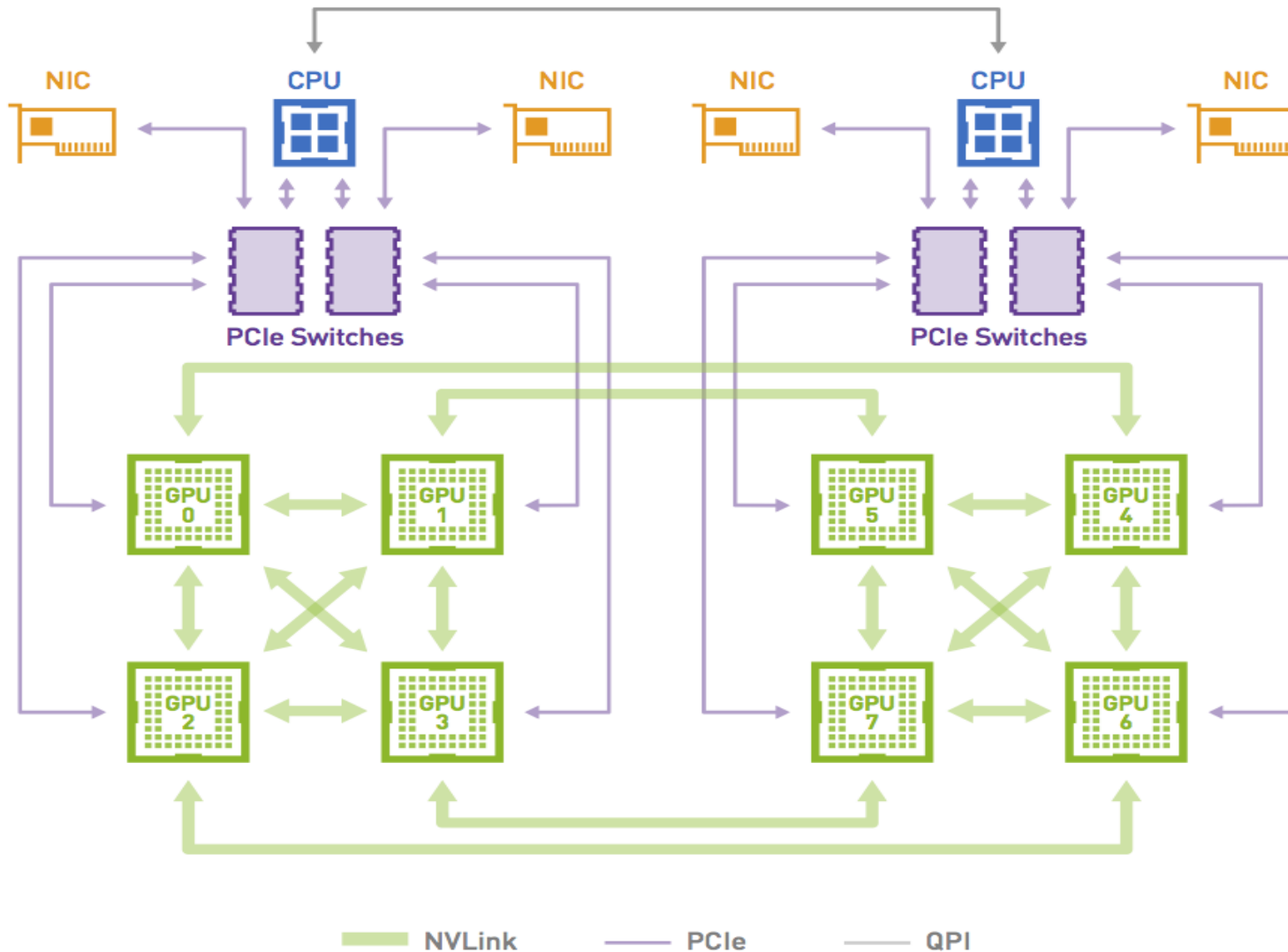
A background network diagram with nodes and connecting lines in shades of gold, brown, and grey.

# AI DEVELOPMENT PLATFORM AS A SERVICE

CENTRE OF EXCELLENCE IN ARTIFICIAL INTELLIGENCE

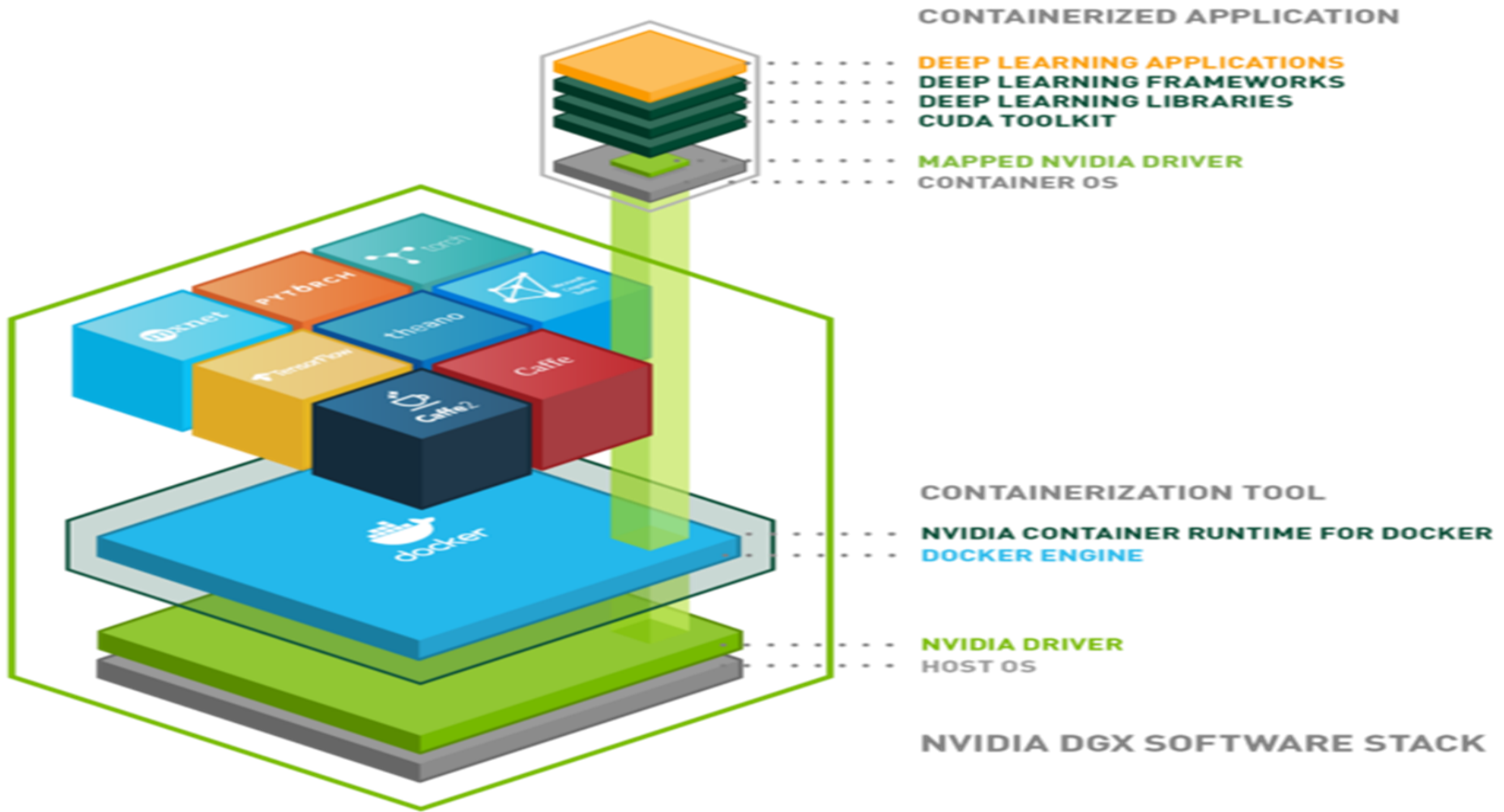
NATIONAL INFORMATICS CENTRE

# SUPER COMPUTING FACILITIES – DGX I SERVERS

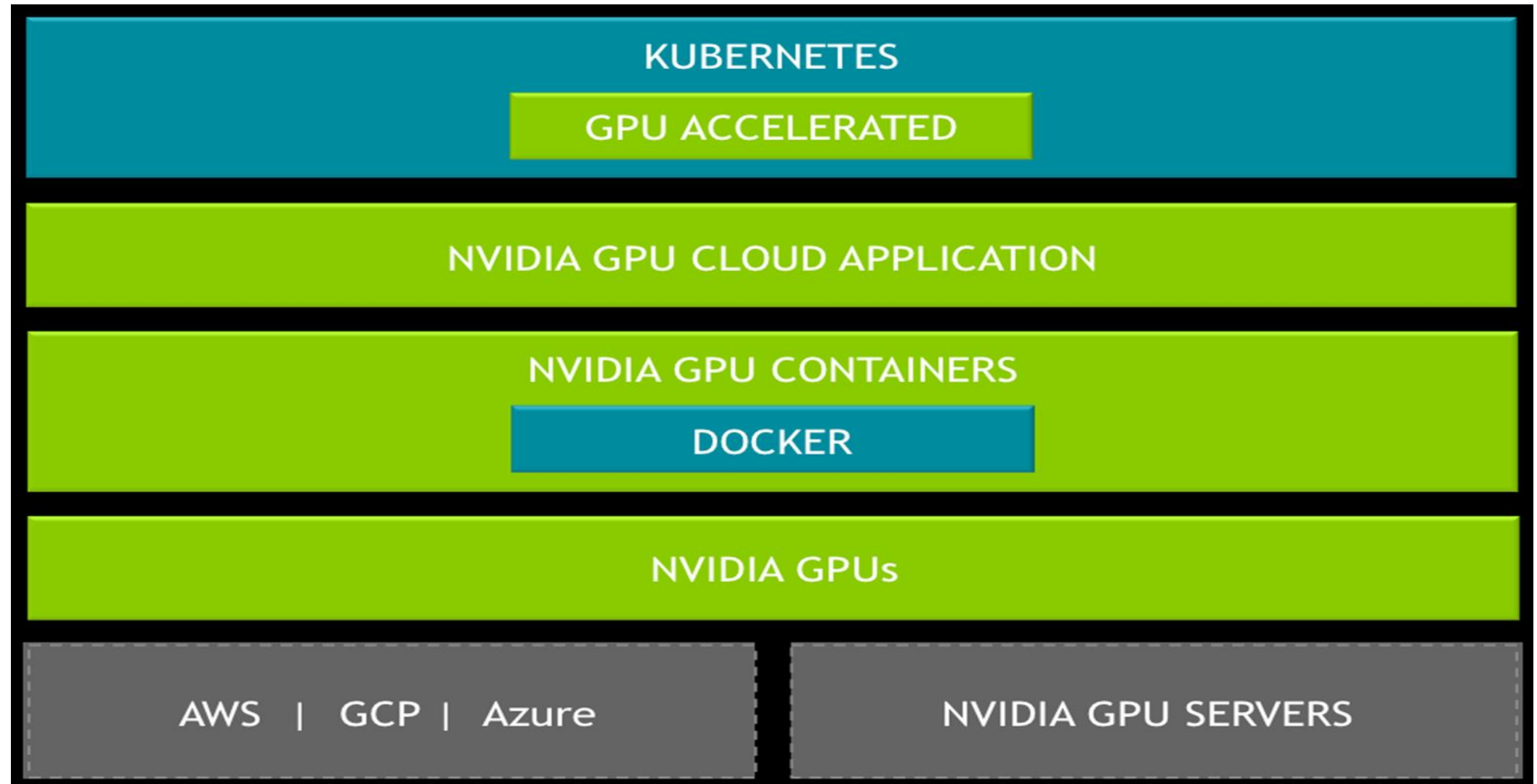


1. NETWORK INTERCONNECT  
4X InfiniBand 100 Gbps EDR  
2X 10 GbE
2. GPUs  
8X NVIDIA Tesla® V100  
32 GB/GPU  
40,960 Total NVIDIA CUDA® Cores  
5,120 Tensor Cores
3. GPU INTERCONNECT  
NVIDIA NVLink™  
Hybrid Cube Mesh
4. SYSTEM MEMORY  
512 GB DDR4 LRDIMM
5. CPUs  
2X 20-Core Intel® Xeon®  
E5-2698 v4 2.2 GHz
6. STREAMING CACHE  
4X 1.92 TB SSDs RAID 0

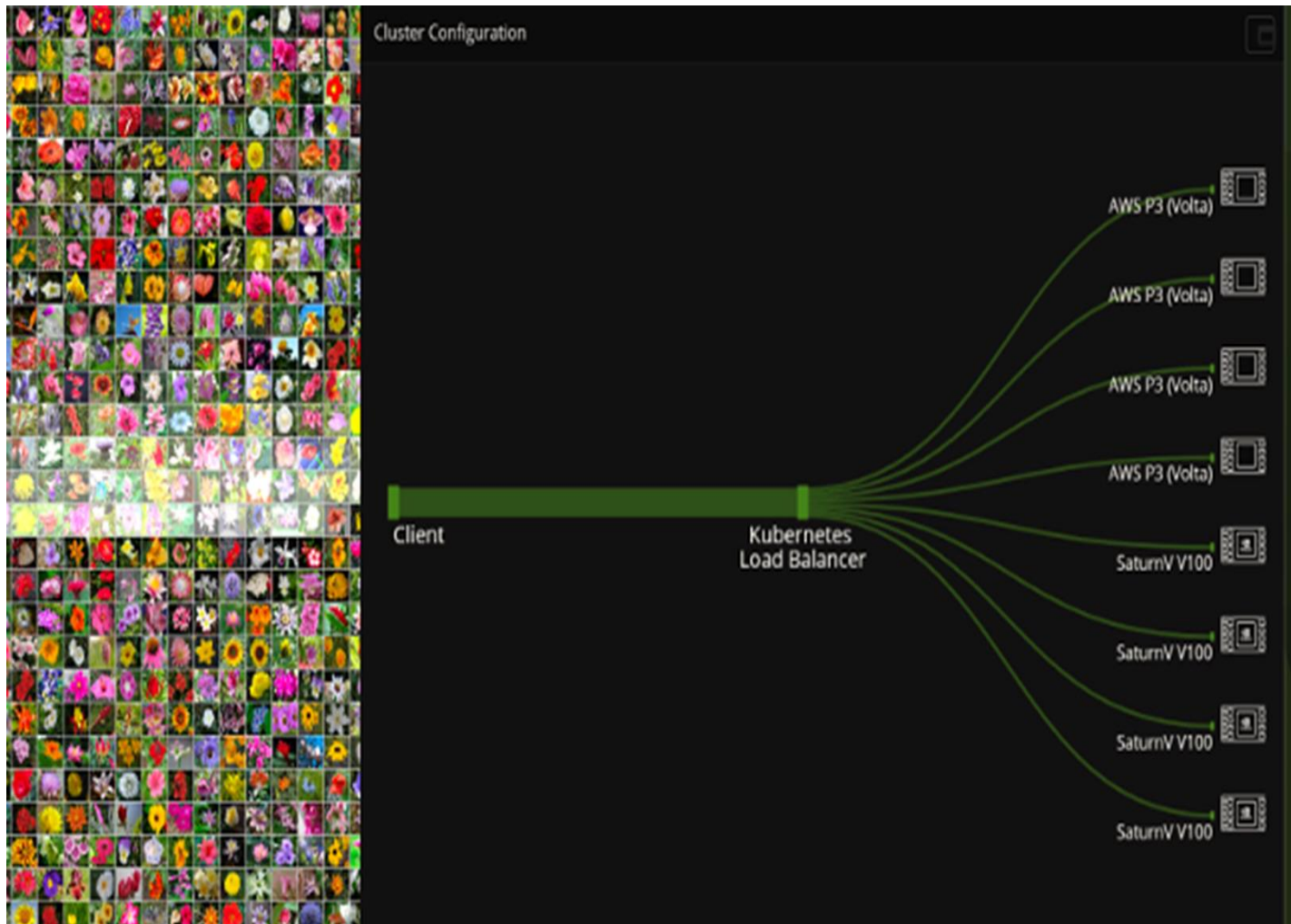
# DEEP LEARNING SOFTWARE STACK FOR THE PLATFORM



# CONTAINER ORCHESTRATION WITH KUBERNETES ACROSS DGX CLUSTER



# AI DEVELOPMENT PLATFORM AS A SERVICE



1. Update State Requirement/ Projects in ai.nic.in under State Initiatives
2. Move eFile for Account – 1 per State
3. Get key & Instructions for Downloading Kubernetes client from COE-AI
4. Prepare Ubuntu based client & Get Ports Opened
5. Get started on the DGX docker container of choice & download & install necessary libraries
6. Train & Test on DGX server in Batches
7. Remove all data loaded on the DGX & copy the Model Back as it is cache memory for computation
8. SSD MEMORY NOT FOR STORAGE
9. Please procure Inferencing Servers for your data centres for production or use cloud VMs in your data centre

# SAMPLE EXAMPLE OF DEEP LEARNING TRAINING & INFERENCE

GPU  
TRAINING  
INFERENCE

## DIGITS FOR DGX-1

A complete GPU-accelerated deep learning workflow

