

Self-Paced Course

Leverage the 'Introduction to AI in the Datacenter' course to learn what AI is and how NVIDIA's GPU technologies and solutions take it to the next level. Designed for enterprise IT professionals and Administrators, this course explores an introduction to AI, GPU computing, NVIDIA AI software architecture, and how to implement and scale AI workloads in the data center. The typical time required to complete the course is 3.5 hours.

[Introduction to AI in the data center](#)

Additional Reading Materials

Please review the links listed below for additional context and details related to AI, GPU computing, NVIDIA AI software, and implementing related workloads in the data center. The typical time required to read these materials is 2 hours.

	Reading time (minutes)		Reading time (minutes)
AI Discover CUDA Zone Learn about the TRITON Inference Server The NGC Catalog What is the difference between AI, ML and DL? What is the difference between DL Training AND Inference? Using GPUs for Robotics Development What is a Recommender system? Accelerating and Enhancing Robotics - ISAAC Robotics Simulation - ISAAC SIM Inference solutions for Video Analytics	1 5 6 6 5 4 5 5 5 5 2	GPU Increase AI Productivity with Multi-Instance GPUs The A100 Tensor Core GPU GPU Communication primitives with NCCL GPT-3 - A 175 billion Parameters Language Model Advanced Multi-GPU communications with NVLINK RTX - Visual computing GPU for workstations	3 4 2 3 4 3
Networking Enhanced data movement with GPUDirect High-speed end-to-end connectivity with DPUs What is a DPU? Cloud scale architecture with DPUs Transforming the data center with DPUs DOCA Software Framework DOCA - Data center infrastructure on a chip	3 4 4 9 5 3 3	Infrastructure Message Passing Interface - MPI What is MLOPS? Deep Learning Frameworks VSPHERE Hypervisor Software Releases NVIDIA AI Enterprise VGPU - Virtualization for IT DCGM Overview DCGM - Data Center GPU Manager	2 7 5 1 3 3 2 2