

GPU Cores and CUDA - Understanding Parallel Processing

CUDA: NVIDIA's parallel computing platform (launched 2007)

GPU Architecture Hierarchy

GPU

Graphics Processing Unit



SMs

Streaming Multiprocessors

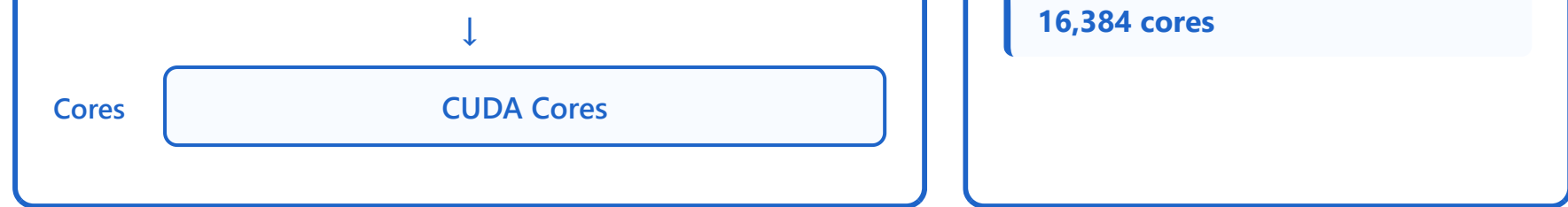
Example

RTX 4090

Streaming Multiprocessors

128 SMs

CUDA Cores



SIMT Execution

Single Instruction, Multiple Threads model for parallel processing

Automatic Integration

PyTorch/TensorFlow automatically leverage CUDA kernels

PyTorch & TensorFlow automatically leverage CUDA for GPU acceleration