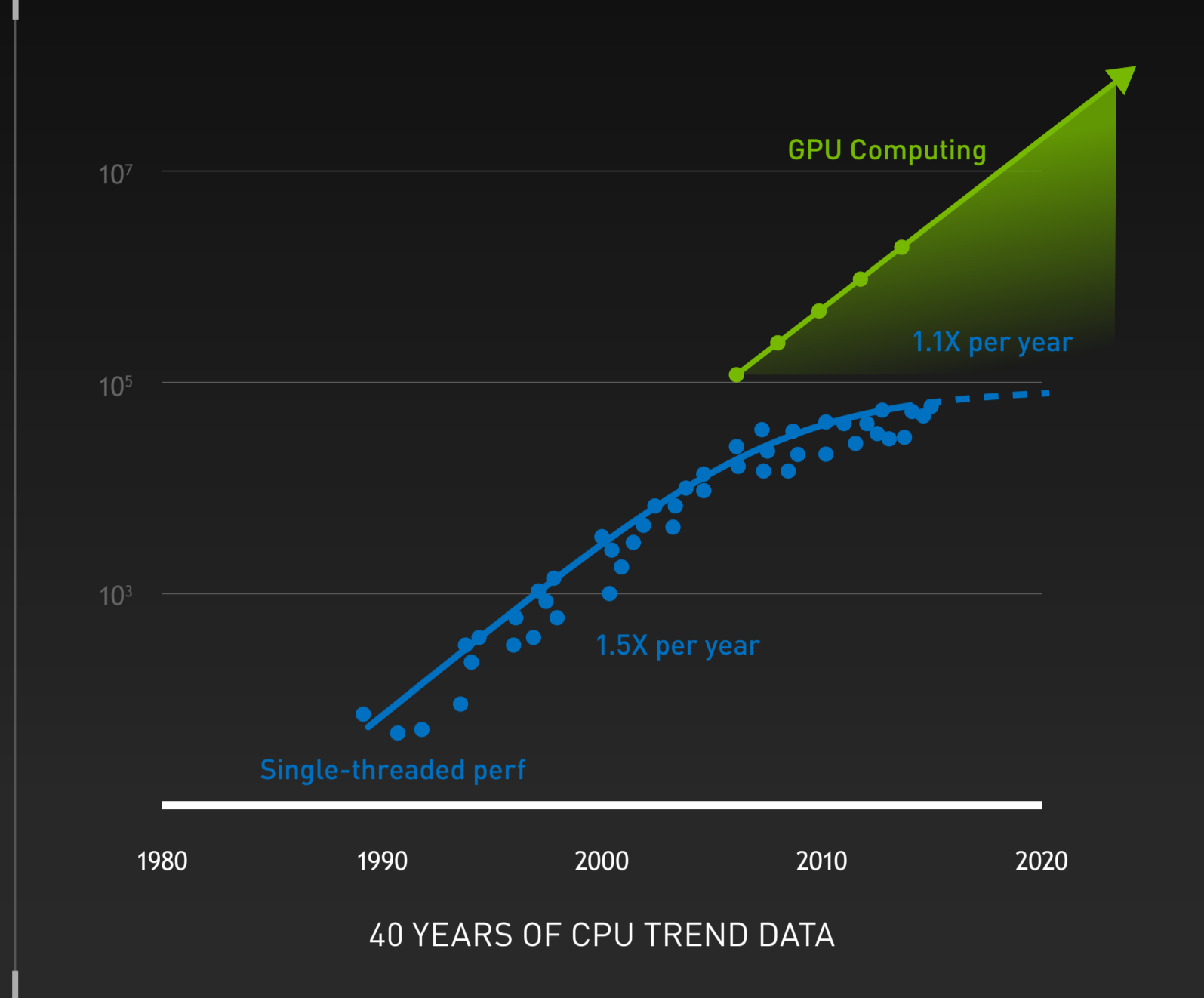


UNLIMITED POSSIBILITIES, EXCEPTIONAL VALUE WITH GPU ACCELERATION



THE RISE OF GPU COMPUTING

As CPU performance levels off, we're seeing the end of Moore's law. And GPU computing is defining a new, supercharged law to replace it.



It starts with a highly specialized Tensor Core GPU and continues through system design, software, algorithms, and optimized applications.

A single GPU-accelerated server replaces dozens of commodity CPU servers, delivering a significant boost in application throughput and cost savings—which means *the more you buy, the more you save*.



SOLVE THE WORLD'S GREATEST CHALLENGES.

With GPU-accelerated deep learning, today's moonshots are within reach.

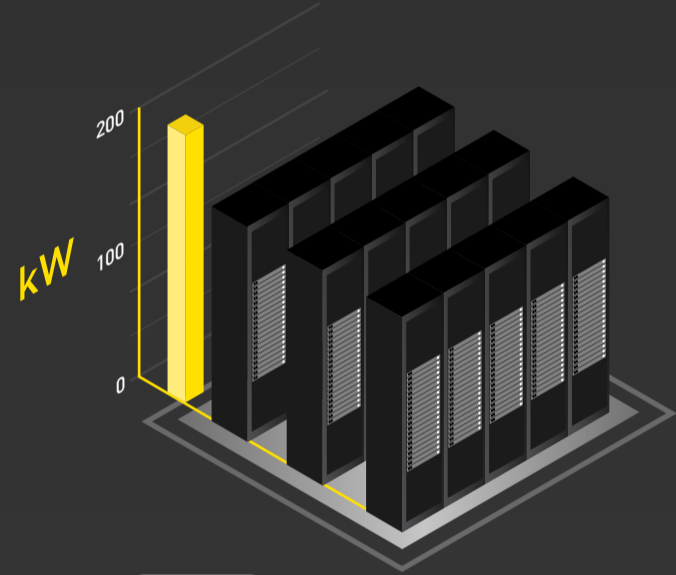
APPLICATION WORKLOAD

DEEP LEARNING TRAINING AND INFERENCE

From medical advancements to creating the next era of autonomous vehicles, deep learning is the key to our future. Implementing NVIDIA GPU acceleration boosts essential training and inference and dramatically reduces server, rack space, and power costs.

CPU Baseline

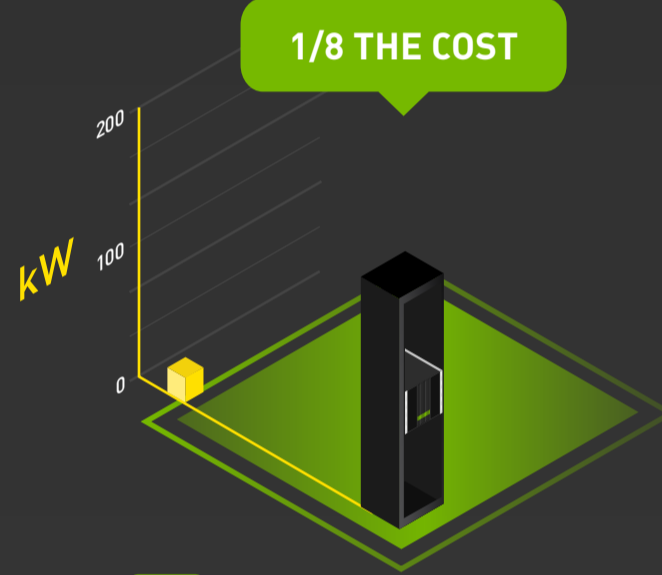
180 Kilowatts



300 Dual CPU servers

GPU Cost Savings

Same Throughput

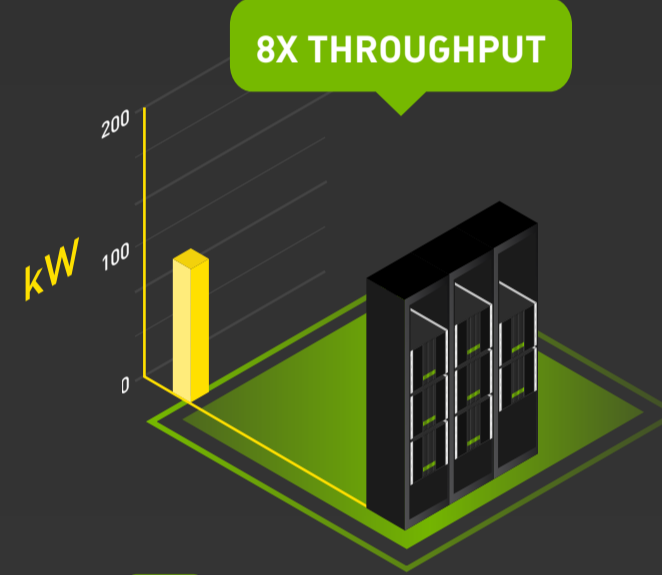


1 HGX-2-based server
with 16 Tesla V100 Tensor Core GPUs

1/8 the power usage | 1/8 the space

GPU Throughput Increase

Time Savings | Same CapEx Budget



8 HGX-2-based servers
with 16 Tesla V100 Tensor Core GPUs each

1/8 the power usage | 1/8 the space

A deep learning application workload includes a balanced system allocation for ResNet-50 training and inference.



BOOST BREAKTHROUGHS IN SCIENCE AND ENGINEERING.

With high performance computing (HPC), our greatest challenges can be our greatest opportunities.

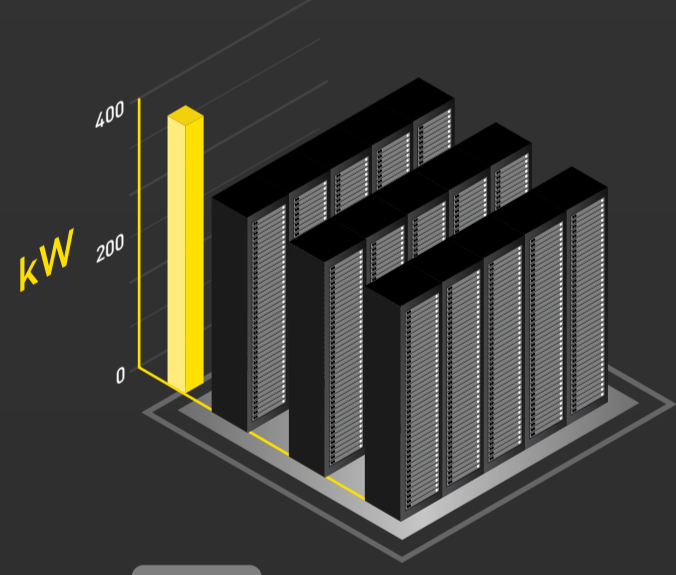
APPLICATION WORKLOAD

HIGH PERFORMANCE COMPUTING

HPC data centers are solving some of the world's most important scientific and engineering challenges. NVIDIA® Tesla® GPUs are the engine of these modern data centers, delivering breakthrough performance with fewer servers, less power consumption, and reduced networking overhead.

CPU Baseline

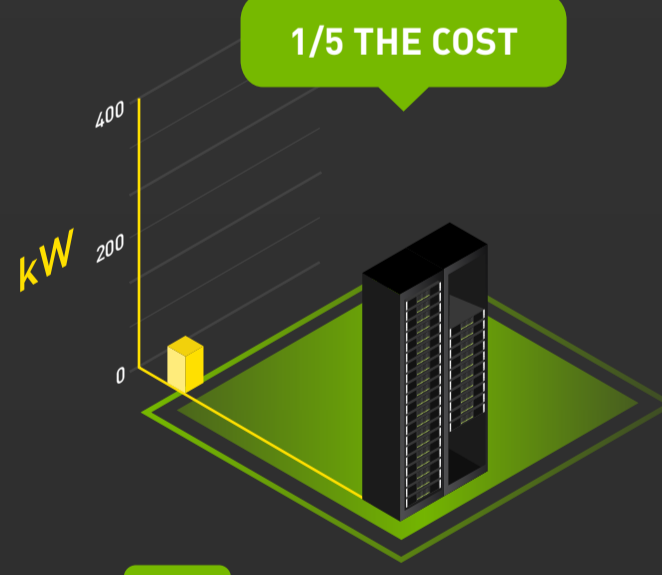
360 Kilowatts



600 Dual CPU servers

GPU Cost Savings

Same Throughput

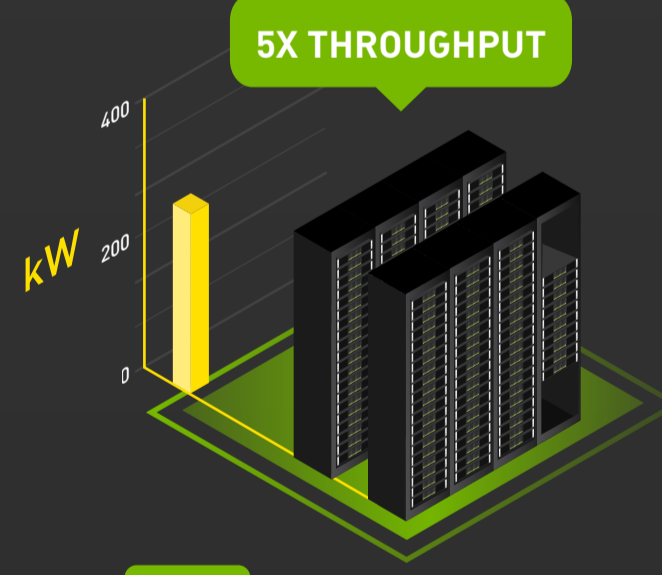


30 Servers
with 4 Tesla V100 Tensor Core GPUs

1/5 the power usage | 1/5 the space

GPU Throughput Increase

Time Savings | Same CapEx Budget



150 Servers
with 4 Tesla V100 Tensor Core GPUs

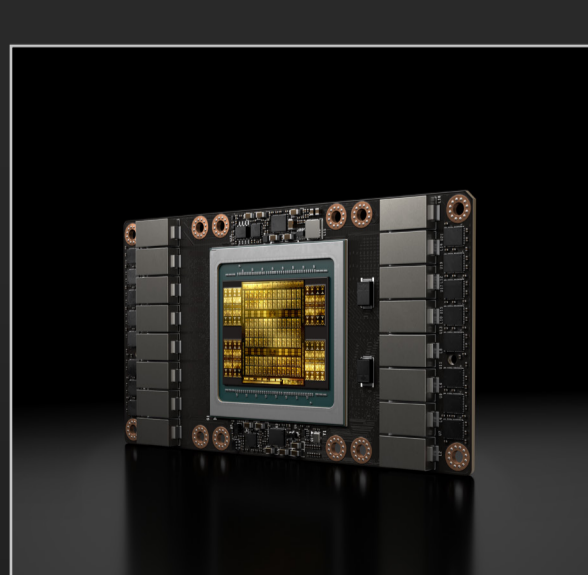
1/5 the power usage | 1/5 the space

A mixed HPC application workload includes an evenly distributed system allocation among key HPC applications such as Amber, Chroma, GTC, LAMMPS, MILC, NAMD, Quantum Espresso, and SPECfem-3D.

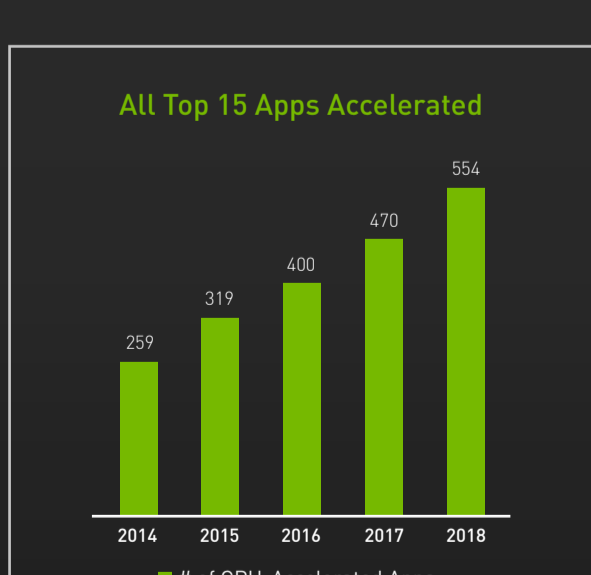


SAVE MONEY ON YOUR HPC AND AI WORKLOADS.

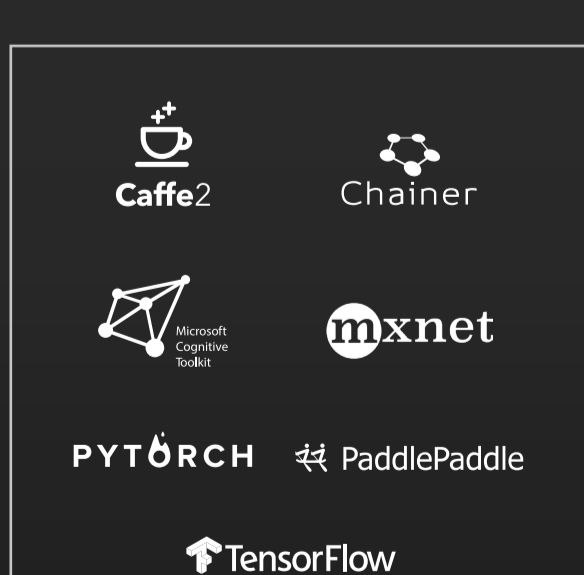
Accelerate Your Scientific and AI Applications
Today with Tesla V100 Tensor Core GPUs.



V100 TENSOR CORE GPU



550+ GPU-ACCELERATED APPLICATIONS



EVERY DEEP LEARNING FRAMEWORK ACCELERATED

The NVIDIA Tesla V100 Tensor Core GPU is the world's most advanced data center accelerator for HPC and AI. It accelerates more than 550 applications—including the top 15 HPC applications and all AI frameworks. From weather prediction and materials science to wind tunnel simulation and genomics, NVIDIA GPU-accelerated computing is at the heart of the most promising areas of discovery.



DISCOVER WHAT'S POSSIBLE—AND HOW MUCH YOU CAN SAVE—WITH GPU ACCELERATION.

www.nvidia.com/v100