

# So why even rent a GPU server for deep learning?

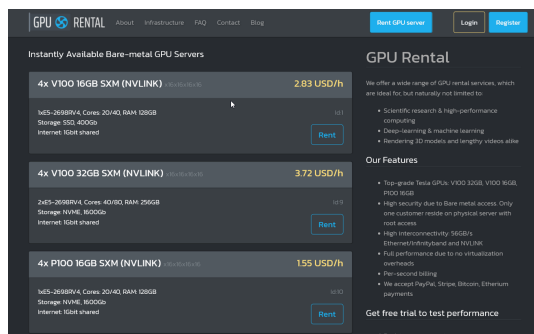
Deep learning can be an ever-accelerating field of machine learning. Major companies like Google, Microsoft, Facebook, and others are now developing their deep learning frameworks with constantly rising complexity and computational size of tasks which are highly optimized for parallel execution on multiple GPU and also multiple GPU servers. So even the most advanced CPU servers are no longer capable of making the critical computation, and this is where GPU server and cluster renting will come in.

Modern Neural Network training, finetuning and 3D rendering calculations usually have different possibilities for parallelisation and could require for processing a GPU cluster (horizontal scaling) or most powerful single GPU server (vertical scaling) and sometime both in complex projects. Rental services enable you to concentrate on your functional <https://gpurental.com/>

scope more instead of managing datacenter, upgrading infra to latest hardware, monitoring of power infra, telecom lines, server health and so on.

## Why are GPUs faster than CPUs anyway?

A typical central processing unit, or a CPU, is a versatile device, capable of handling many different tasks with limited parallelism certainly using tens of CPU cores. A graphical digesting unit, or perhaps a GPU, was created with a specific goal in mind - to render graphics as fast as possible, which means doing a large amount of floating point computations with huge parallelism making use of a large number of tiny GPU cores. This is why, thanks to a deliberately large amount of specialized and sophisticated optimizations, GPUs tend to run faster than traditional CPUs for particular tasks like Matrix multiplication that is clearly a base task for Deep Learning or 3D Rendering.



The screenshot displays the GPU RENTAL website interface. The header includes the logo and navigation links: About, Infrastructure, FAQ, Contact, Blog. There are buttons for 'Rent GPU server', 'Login', and 'Register'. The main content area is titled 'Instantly Available Bare-metal GPU Servers' and lists three server configurations:

Configuration	Price (USD/h)
4x V100 16GB 5XM (NVLINK) Intel Xeon E5-2680v4, 20x140 GB RAM, 128GB SSD, 400Gb/s Internet 10Gbit shared	2.83
4x V100 32GB 5XM (NVLINK) Intel Xeon E5-2680v4, 20x140 GB RAM, 256GB SSD, 400Gb/s Internet 10Gbit shared	3.72
4x P100 16GB 5XM (NVLINK) Intel Xeon E5-2680v4, 20x140 GB RAM, 128GB SSD, 400Gb/s Internet 10Gbit shared	1.55

Each configuration has a 'Rent' button. To the right, the 'GPU Rental' section explains the service and lists features:

- Scientific research & high-performance computing
- Deep learning & machine learning
- Rendering 3D models and lengthy videos
- Top-grade Tesla GPUs: V100 32GB, V100 16GB, P100 16GB
- High security due to Bare metal access. Only one customer reside on physical server with root access
- High interconnectivity: 56GB/s Ethernet InfiniBand and NVLink
- Full performance due to no virtualization overheads
- Pay-as-you-go billing
- We accept PayPal, Stripe, Bitcoin, Ethereum payments

At the bottom, it says 'Get free trial to test performance' and '1 Register new account'.