VULTR

How to speed time-to-market with new visual computing and ML initiatives with cloud GPUs

The cloud GPU advantage

Using cloud-based GPUs is a cost-effective solution because it allows clients to only pay for the resources they need instead of purchasing expensive GPUs and installing them locally. Cloud computing models are becoming increasingly popular, especially due to the recent shift toward remote working. Leveraging cloud computing models, including cloud-based GPUs, provides numerous benefits for organizations, such as:

- Eliminating procurement-related bottlenecks, thereby gaining time-to-market.
- Saving on GPU hardware and maintenance costs.
- Cost-effective global scaling of a GPU's computational power up or down as needed.
- Lowering the cost of digital operation for new AI and ML initiatives to free resources for digital innovation.
- Improving productivity and innovation by giving organizations access to the latest GPU technology when using cloud GPUs, as cloud providers are better at acquiring and installing the latest technologies than small and medium-sized companies. Moreover, cloud GPU clients have the chance to finish their heavy work requiring powerful GPU processing quickly. This is because cloud GPU resources can be scaled easily compared with local GPUs, improving the company's overall productivity.

Cloud GPUs are a recent advancement in the cloud computing sphere. The GPU is a microprocessor chip used to accelerate graphics rendering, but unlike CPUs, they can process data in parallel. This makes them very useful for intensive processing tasks, such as:

- Completing regular tasks like video editing and gaming.
- Carrying out data analytics, such as with matrix multiplication, in which a GPU can process multiple data using a single instruction. This is also known as single instruction with multiple data (SIMD).
- Improving the speed of training machine learning models because all operations can be run simultaneously by a GPU instead of running them one after another.

Cloud-based GPUs enable clients to access advanced GPU technology without needing to deploy it locally on end-user devices while still paying a use-based subscription fee. Cloud GPUs also remove the need for regular updates and maintenance offloading these tasks to the provider.

Benefits of cloud GPUs

With cloud-based GPUs, your organization can use the latest GPU technology on demand and conduct advanced computing operations while eliminating the associated costs of purchasing, installing, and maintaining GPUs on local infrastructure. The next section will list more advantages of using cloud GPUs over on-premise alternatives.

Time-to-market

Cloud GPUs can be set up and configured quickly, giving your projects the right-sized GPU resources using cost-effective payment options. Furthermore, there's an ongoing global shortage of semiconductors as the demand outstrips supply, with manufacturers struggling to keep up. By using cloud GPUs, you can avoid the hassle of procuring and installing GPUs and the obstacle of supply chain shortages.

Freeing up resources to accelerate digital innovation

There's no upfront investment with cloud GPUs – you only pay for what you use. For example, the Vultr Cloud GPU, powered by NVIDIA, provides a cost-effective cloud GPU solution that starts as low as $90 per month.

Scalability

As organizations grow in business and workload, they must scale their IT infrastructure to match. Scaling on-premises IT infrastructure is costly and may not always be feasible. For example, your organization could be working on a specific project that requires considerable GPU processing power for three months. Installing local GPUs only wastes resources and isn't a viable solution for such a short time.

Meanwhile, cloud-based GPUs grow to meet demand without needing any additional infrastructure. For instance, if you need more computational power to perform data analysis, you can instantly increase the cloud GPU resources without purchasing anything. Fractional cloud GPUs lower the costs of getting started because there's no need to buy an entire GPU. You can only rent a fraction of one GPU and easily save 50 percent of the cost, or maybe even more, compared to a whole GPU.

Adaptability

When utilizing a cloud GPU, you can experiment with different GPUs until you ultimately find the one that best fits your needs.

For example, GPUs such as the NVIDIA A100 Tensor Core GPU are best suited for training deep learning (DL) models, while others, such as the NVIDIA A 40, are best for visual effects (VFX), rendering, and virtual workstations.

On the other hand, updating existing GPUs is easier when hosted in the cloud rather than purchasing new chips and installing them on on-premise computers. Cloud service providers typically adopt new GPUs, such as the NVIDIA H100, when they become available.

Agility

Installing GPUs on local computers consumes space and electricity, and they require continuous maintenance and other climate conditions, such as cooling and heating. Besides, local computers slow down when running DL models or processing intensive data analytics tasks. This prevents you from using these devices for other jobs, lowering overall performance and decreasing work throughput.

You don't need to worry about these issues when outsourcing your GPU workloads to a cloud provider. Outsourcing GPU workloads makes you more agile because you free up local computing resources by offloading work to the cloud. Therefore, you can do more locally and enjoy an increased performance output.

Availability

The primary motivation behind cloud computing models is to allow remote access. Cloud GPU platforms like Vultr have a worldwide footprint, and you can access them from anywhere through the Internet.

Why choose Vultr Cloud GPUs?

Vultr Cloud GPU, powered by NVIDIA, is accessible worldwide. The following features distinguish Vultr Cloud GPU from others:

- Vultr allows customers to rent fractional GPUs, meaning you don't need to rent the entire GPU, just the virtual partition you need to perform your desired tasks. When the time comes to scale, you can move to bare metal servers with multiple GPUs. This can save your organization a considerable amount of money over time.
- Vultr has an advanced platform with worldwide access to its GPU technology. Vultr also powers its platform with a web firewall and distributed denial of service protection to prevent cyberattacks from disrupting your work when using cloud GPU infrastructure remotely.
- Vultr provides different GPUs, such as the NVIDIA A100 and A40, each able to tackle a specific workload more efficiently. The unique feature of Vultr is that it allows customers to change GPU types on the fly and upgrade existing GPUs quickly to meet their changing workload demands.

Conclusion

The cloud computing market is expanding rapidly and has been integrated into most organizations' IT environments. The growth of emerging technologies like DL, AI, and machine learning has made the need for cloud GPUs urgent.

Cloud GPUs offer unpreceded value in terms of cost, performance, and total cost of ownership compared with local GPUs. Vultr provides a unique cloud GPU offering that allows all organizations, regardless of their size and industry, to benefit from the cutting-edge technology of GPU quickly and affordably.

You can learn more about Vultr Cloud GPU by visiting the product page.