Competitive advantage means making more accurate decisions, faster. Unfortunately, the massive volume of data that has to be analyzed to extract that sort of value has traditionally required long compute cycles and a physically large—and costly—computing infrastructure.

However, advances in graphics processing unit (GPU)-accelerated computing mean you can now take computing challenges once considered impossible to overcome and transform them into immediate financial gain.

**What is GPU-accelerated computing?**

GPU-accelerated computing occurs when you use a GPU in combination with a central processing unit (CPU), letting the GPU handle as much of the parallel process application code as possible. The GPU takes the parallel computing approach orders of magnitude beyond the CPU, offering thousands of compute cores. This can accelerate some software by 100x over a CPU alone. Plus, the GPU achieves this acceleration while being more power- and cost-efficient than a CPU.

Provided your system design team is experienced with building both CPU and GPU-based systems and the storage subsystems required for this level of data analytics, the outcome of moving to a GPU-accelerated strategy is superior performance by all measures, faster compute time, and reduced hardware requirements.

**Benefits of GPU-accelerated computing**

- **Computing Power/Speed** — A single GPU can offer the performance of hundreds of CPUs for certain workloads. In fact, NVIDIA®, a leading GPU developer, predicts that GPUs will help provide a 1000X acceleration in compute performance by 2025.

- **Efficiency/Cost** — Adding a single GPU-accelerated server costs much less in upfront, capital expenses and, because less equipment is required, reduces footprint and operational costs. Using libraries also allows organizations to use GPU acceleration without in-depth knowledge of GPU programming, reducing the investment of time required to achieve results.

- **Flexibility** — The inherently flexible nature of GPU programmability allows new algorithms to be developed and deployed quickly across a variety of industries. According to Intersect360 Research, 70% of the most popular HPC applications, including 10 of the top 10, have built-in support for GPUs.

- **Long-term Benefits** — Adding GPU-accelerated computing now prepares you for the artificial intelligence (AI) revolution, which also relies in GPU-accelerated computing. This inevitable increase on the reliance on GPUs means that early adopters will enjoy not only greater computing power over time but have a greater margin of difference over time than competitors who do not migrate to GPU-accelerated computing.
Popular GPU-accelerated options from Penguin Computing include:

Relion XO1114GTS
- OpenCompute form factor
- Up to four (4) Nvidia V100 SXM2 with NVLink for low latency Peer-2-Peer compute
- Delivers speed and efficiency with two Intel® Xeon® Scalable Processors and up to 3TB DDR4-2666MHz ECC memory
- Single root complex with four (4) Nvidia V100 SXM2 to a single CPU with GPU-Direct support for low latency interconnect
- Up to four (4) 2.5" hot-swap drive bays with SAS/SATA support and SATA DOM for a range of direct storage options
- Adheres to OCP specifications and OCP racks to minimize operational costs while maximizing performance

Relion XE2118GT
- EIA form factor
- Up to eight (8) PCIe GPUs or coprocessors for high-density performance
- Delivers speed and efficiency with two Intel® Xeon® Scalable Processors and up to 3TB DDR4-2666MHz ECC memory
- Dual root complex with four (4) GPUs or coprocessors from each CPU for a balanced compute configuration
- Up to eight 2.5" hot-swap drive bays with SAS/SATA support and SATA DOM for a range of direct storage options

Learn More
To learn more about GPU-accelerated computing visit: www.penguincomputing.com/gpu

Purchase with Financing
Finance products, services, even soft costs with Penguin Computing Capital. Choose from options such as no money down, flexible billing choices, extended repayment timelines, and a variety of end-of-term alternatives.

About Penguin Computing
Penguin Computing, Inc. is a 20-year-old, U.S.-based global provider of high-performance computing (HPC), artificial intelligence (AI), and data center solutions with more than 2,500 customers in 40 countries, across eight major vertical markets. Penguin Computing offers a comprehensive portfolio of hardware, software, and services, including solutions based on the Open Compute Project (OCP), as well as financing and top-rated customer support. Penguin Computing products include Linux-based servers, software, integrated, turn-key clusters, enterprise-grade storage, and bare metal HPC on cloud via Penguin Computing® On-Demand™ (POD).