



NVIDIA Dramatically Simplifies Parallel Programming With CUDA 6

Unified Memory, Drop-In Libraries Among New Programmability Features to Empower Next Wave of GPU Developers

SANTA CLARA, CA -- NVIDIA today announced NVIDIA® CUDA® 6, the latest version of the world's most pervasive parallel computing platform and programming model.

The CUDA 6 platform makes parallel programming easier than ever, enabling software developers to dramatically decrease the time and effort required to accelerate their scientific, engineering, enterprise and other applications with GPUs.

It offers new performance enhancements that enable developers to instantly accelerate applications up to 8X by simply replacing existing CPU-based libraries. Key features of CUDA 6 include:

- **Unified Memory** -- Simplifies programming by enabling applications to access CPU and GPU memory without the need to manually copy data from one to the other, and makes it easier to add support for GPU acceleration in a wide range of programming languages.
- **Drop-in Libraries** -- Automatically accelerates applications' BLAS and FFTW calculations by up to 8X by simply replacing the existing CPU libraries with the GPU-accelerated equivalents.
- **Multi-GPU Scaling** -- Re-designed BLAS and FFT GPU libraries automatically scale performance across up to eight GPUs in a single node, delivering over nine teraflops of double precision performance per node, and supporting larger workloads than ever before (up to 512GB). Multi-GPU scaling can also be used with the new BLAS drop-in library.

"By automatically handling data management, Unified Memory enables us to quickly prototype kernels running on the GPU and reduces code complexity, cutting development time by up to 50 percent," said Rob Hoekstra, manager of Scalable Algorithms Department at Sandia National Laboratories. "Having this capability will be very useful as we determine future programming model choices and port more sophisticated, larger codes to GPUs."

"Our technologies have helped major studios, game developers and animators create visually stunning 3D animations and effects," said Paul Doyle, CEO at Fabric Engine, Inc. "They have been urging us to add support for acceleration on NVIDIA GPUs, but memory management proved too difficult a challenge when dealing with the complex use cases in production. With Unified Memory, this is handled automatically, allowing the Fabric compiler to target NVIDIA GPUs and enabling our customers to run their applications up to 10X faster."

In addition to the new features, the CUDA 6 platform offers a full suite of programming tools, GPU-accelerated math libraries, documentation and programming guides.

Version 6 of the CUDA Toolkit is expected to be available in early 2014. Members of the CUDA-GPU Computing Registered Developer Program will be notified when it is available for download. To join the program, [register here](#).

For more information about the CUDA 6 platform, visit NVIDIA booth 613 at [SC13](#), Nov. 18-21 in Denver, and the [NVIDIA CUDA website](#).

About CUDA

CUDA is a [parallel computing platform and programming model](#) developed by NVIDIA. It enables dramatic increases in computing performance by harnessing the power of GPUs. With more than two million downloads, supporting more than 240 leading engineering, scientific and commercial applications, the CUDA programming model is taught in [over 700 universities and institutions](#) worldwide and is the most popular way for developers to take advantage of GPU-accelerated computing.

To Keep Current on NVIDIA:

- Like NVIDIA on [Facebook](#).
- Connect with NVIDIA on [LinkedIn](#).
- Follow [@NVIDIATesla](#) and [@GPUComputing](#) on Twitter.
- View NVIDIA videos on [YouTube](#).
- Keep up with the [NVIDIA Blog](#) and [Parallel ForAll developer blog](#).
- Use the Pulse news reader to subscribe to the NVIDIA Daily News feed.

Ken Brown
Corporate Communications
+1-408-486-2626

