Hardware accelerators are becoming increasingly important in the HPC community. Our performance monitor VampirTrace provides CUDA and OpenCL support to give detailed insight into the runtime behavior of hardware accelerators. This enables extensive performance analysis and optimization of hybrid programs.

**CUDA and OpenCL**

The Compute Unified Device Architecture (CUDA) is NVIDIA's approach to providing access to graphics accelerators for general purpose computing (GPGPU). CUDA allows scientific applications to take full advantage of NVIDIA's accelerator capabilities.

The Open Computing Language (OpenCL) is a programming standard for parallel computing on heterogeneous platforms. The specification is more abstract than hardware specific APIs, and is developed to be portable between different platforms that support OpenCL.

**Accelerator Monitoring**

VampirTrace captures accelerator API calls with its generic library tracing capability. More details are obtained by interpreting the functions' semantic and extending their VampirTrace wrappers. This includes timing information of fine-grained parallel functions (kernels) and data transfers between host and accelerator devices.

Using the preload mechanism of a dynamic linker traces can be easily generated without the need for recompilation. Optional customization via the runtime environment variables allows the user to control the event recording. Performance counters on accelerator devices can be captured with NVIDIA's CUPTI library.

**Key Features**

- Capture GPU idle time
- Measurement of kernel execution times and data transfers between host and accelerator
- Record GPU memory usage and detect memory leaks
- No recompilation required
- Customizable via environment variables
- CUDA ≥ 2.3 incl. CUPTI and OpenCL ≥ 1.0 support
- Tracing overhead typically < 1%

---

**Contact**

Technische Universität Dresden  
Center for Information Services and HPC (ZIH)  
01062 Dresden, Germany  
E-mail: zih@tu-dresden.de  
Web: www.tu-dresden.de/zih/vampirtrace