

NVIDIA CUDA™

Lucas Vaz Pires

Mapa da apresentação

- Funcionamento e Generalidades
- Linguagens Suportadas
- Modelo de Programação
- Exemplos de aplicações
- Produtos no Mercado

Funcionamento e Generalidades

- CUDA™ é o acrônimo de *Compute Unified Device Architecture*.
- Paradigma de computação paralela proposto pela empresa NVIDIA.
- Propõem o uso de unidades de processamento gráfico para uso geral.

Funcionamento e Generalidades

- Tecnologia potencialmente barata.
- Otimização das placas mais recentes (arquiteturalmente).

Linguagens Suportadas

- Diversas são as linguagens suportadas pela tecnologia, entretanto, algumas como: C, C++ e Fortran proveem facilidades como enviar códigos diretamente a GPU, sem necessitar de compilação.
- *CUDA™ ISA (Instruction Set Architecture).*

Modelo de Programação

- Plataforma de computação paralela .
- Exige que o usuário configure todas as informações inerentes a uso de memórias de acesso global, caches, threads, etc.

Modelo de Programação

- Paradigma de *computação com GPU*, onde os aplicativos podem ser acelerados por GPU, sob processamento em paralelo, e as partes sequenciais de suas cargas de trabalho na CPU.

Modelo de Programação

- Pode acontecer *overheads*, por conta tanto da aplicação, por não ser tão adequada ao paradigma, quanto pelo uso de um barramento desfavorável ao todo do sistema.
- O software `CUDA™` apenas suporta equipamentos NVIDIA.

Exemplos de aplicações

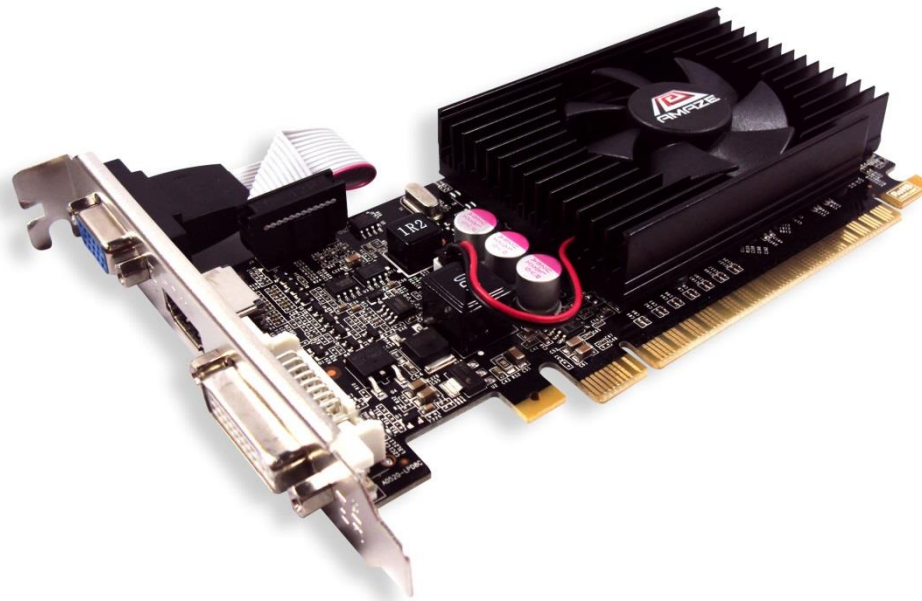
- *Alenka*: aplicação baseada em SQL para armazenagem e processamento de dados em hardwares CUDA. É uma aplicação acadêmica.
- *AgiSoft PhotoScan*: é uma solução para reconstituição de imagens com partes perdidas, a partir de informações 3D iniciais. É uma aplicação comercial.

Produtos no Mercado

- GeForce GT 610
 - Núcleos CUDA: 48
 - Velocidade de clock do processador: 1620
 - Velocidade de clock de gráficos (Mhz):
810
 - Velocidade da memória: 1.8 Gbps
 - Custo aproximado: \$100,00

Produtos no Mercado

- GeForce GT 610



Produtos no Mercado

- Tesla K20X
 - Pico de performance de ponto flutuante de precisão dupla: 1.31 Tflops
 - Pico de performance de ponto flutuante de precisão simples: 3.95 Tflops
 - Largura de banda de memória (ECC desligado): 250 GB/sec
 - Tamanho da memória (GDDR5): 6 GB
 - Núcleos CUDA: 2688
 - Custo aproximado: \$3.800,00

Produtos no Mercado

- Tesla K20X

