

NVIDIA e o futuro do GPGPU



Richard Cameron
Country Manager da
NVIDIA no Brasil

A NVIDIA é uma das empresas que mais tem se esforçado para a migração do processamento para plataformas híbridas com GPGPU.

Conversamos com Richard Cameron, Country Manager da NVIDIA no Brasil, sobre os futuros lançamentos desta empresa para o mercado de GPGPU e também quanto à sua aceitação no Brasil.

Grandes mudanças não são bem-vistas, principalmente quando falamos de Informática. Isto porque deve haver um estudo de compatibilidade entre o novo que chega e todo o sistema existente, análises de risco e impacto, entre outros procedimentos.

O GPGPU, apesar de ser uma evolução natural para os nossos computadores, passa por esta mesma dificuldade. Enquanto determinadas áreas que sofriam um enorme "gargalo" de processamento estão entusiasmadas com as mudanças prestes a acontecer, outras continuam apáticas às mudanças, ou mesmo as desconhecem.

Conversamos com o Country Manager da NVIDIA Brasil, Richard Cameron, sobre o mercado brasileiro e GPGPU.

SE: A NVIDIA desde 2007 oferece soluções para o GPGPU com o CUDA e, depois, com sua plataforma para supercomputação chamada Tesla. Agora em 2010 ela lançará uma nova família de GPUs, conhecida como Fermi. Quais serão as novidades desta geração?

O Fermi é uma GPU que foi pensada desde o princípio, na sua arquitetura, para aplicações em supercomputação. Recursos como memórias com ECC e *double point precision* o fazem ideal para aplicações que exigem precisão muito grande, como *crash tests* ou túnel do vento, na indústria automobilística, pois os resultados destes testes podem significar a vida ou a morte de pessoas.

Além disso, a próxima versão do CUDA dará suporte a um número maior de linguagens de alto nível, como o C++, além de integração com o Microsoft Visual Studio, facilitando ainda mais o desenvolvimento de aplicativos para nossa base.

E como está a disponibilidade da plataforma Tesla no Brasil?

O Tesla é vendido na forma de placa ou de servidores. Uma empresa brasileira já está fechando acordos para a fabricação de servidores com ele no Brasil. Adicionalmente, existem outras empresas, por exemplo, que estão fazendo servidores completos. Elas compram nossas lâminas de Tesla e montam racks para vender um servidor completo, como a Bull fez com a Petrobras.

Com os supercomputadores híbridos nascendo com custos acessíveis, o que a NVIDIA acredita que acontecerá com os nossos computadores baseados no x86. Deixarão de existir?

O x86 continuará existindo, há muitas coisas que continuarão fazendo da CPU um processador melhor, mesmo com o avanço de GPGPU. Ninguém tem a intenção de paralelizar o processador numa GPU, porque ele não vai ficar mais rápido. ▶

Mas o que nós veremos acontecer a longo prazo é a figura do computador mudar em relação ao que sempre foi: aquela arquitetura com volume de armazenamento muito grande e próximo do usuário. Agora a computação pessoal vai virar de ponta-cabeça e irá para longe do usuário. Inclusive, existem vários aplicativos baseados em Tegra que estão em Cloud Computing.

Cloud Computing, então, é uma verdade inevitável?

O fundador da Sun, Scott McNealy, disse há trinta anos uma frase que tornou-se uma profecia: "The computer is the network" ("O Computador é a rede", em tradução livre). Hoje, um computador sozinho não é nada, mas a partir do momento que ele se conecta à internet, ele produz conhecimento.

A computação está migrando para a nuvem, os nossos computadores serão apenas uma interface de acesso. Todo o poder de processamento que o usuário precisa para o processamento de textos, desenvolvimento, jogos, etc será dado por grandes datacenters conectados a rede. Se Scott McNealy tivesse dito aquela frase neste momento, ela seria: "The computer is the cloud" ("O computador é a nuvem", em tradução livre).

Quais áreas estão sendo mais receptivas às mudanças?

Em primeiro lugar as áreas que têm enormes volumes de dados para serem processados, principalmente aqueles voltados para a pesquisa científica. Empresas dos ramos petrolífero e automobilístico criaram uma grande demanda por produtos que usem GPGPU.

Um exemplo que podemos citar é a Universidade Federal de Santa Catarina, os estudantes de informática, junto com o laboratório de Medicina, desenvolveram softwares para análise de tomografias computadorizadas que roda na arquitetura Tesla.

Já que as áreas científicas têm sido as mais receptivas, como a NVIDIA encara o GPGPU para o desenvolvimento de um país?

A GPU consiste no processador mais poderoso e acessível do planeta com centenas de aplicações já desenvolvidas pra ele, e hoje é o processador usado para salvar vidas, extrair petróleo, para desenvolver carros mais seguros e aviões mais econômicos. A

“As GPU está relacionada ao desenvolvimento de um país, na economia, geração de riquezas, geração de conhecimento, etc, porque ela faz o que a CPU faz, mas muitas vezes mais rapidamente, barato e ecologicamente correto. É a quebra de paradigma de processamento.”

GPU está relacionada ao desenvolvimento de um país, na economia, geração de riquezas, geração de conhecimento, etc, porque ela faz o que a CPU faz, mas muitas vezes mais rapidamente, barato e ecologicamente correto. É uma quebra de paradigma de processamento.

Recentemente aconteceu uma mudança na classificação fiscal das placas de vídeo, para incentivo da produção local no Brasil. A NVIDIA acha que estas mudanças são suficientes para começarmos a ter bons produtos fabricados no Brasil?

Por questões de ciclo de vida, a placa de vídeo é o componente mais complicado e arriscado de se produzir no Brasil, pois você pega um computador e analisa todos os componentes, ela é o que tem o menor ciclo de vida, em média de 6 a 9 meses.

Um fabricante brasileiro leva, em média, 3 meses para analisar um novo produto que é lançado lá fora. Depois de fazer um pedido do produto, leva-se mais 2 meses para as peças chegarem no Brasil e depois mais 1 mês para fabricar. Nestes 6 meses o produto já está em fase de declínio no seu ciclo de vida.

Também com relação ao ciclo de vida, o fabricante tem algo chamado de NRE, Non-Recurring Expense, que são custos fixos que precisarão ser diluídos ao longo de no mínimo 6 meses para recuperar o investimento. Mas não existem 6 meses para diluir o investimento na produção de placas de vídeo no Brasil.

É por isso que quem produz placas de vídeo no Brasil, faz as placas mais básicas e diz que o mercado terá que se contentar com isso, o que não é justo para o consumidor.

Mas o PPB não ajuda como incentivo para produção de placas voltadas para GPGPU no Brasil?

Na verdade, aqui existe um segundo problema. O PPB classifica as placas de vídeo como um item de luxo, apenas para jogos. Esta lei define que um integrador ou fabricante local pode colocar uma placa de vídeo até o limite de 10% das suas máquinas, mas dentro dessa porcentagem ele precisa escolher entre outras peças.

Com este limite, e os riscos já comentados para a fabricação de placas localmente, é impossível fabricar, no país, placas de vídeo com bom desempenho para o GPGPU. **E**