

Stručni kurs: Ispod haube GPU programiranja

1. Cilj kursa

Cilj kursa je da studenti razumeju šta se događa „ispod haube“ pri izvršavanju programa na grafičkim procesorima: kako se programski model preslikava na hardver i kakvu ulogu imaju prevodilac, runtime sistem, drajver, memorijska hijerarhija i alati za praćenje izvršavanja.

2. Očekivana predznanja

- Osnovno poznavanje jezika C ili C++
- Osnovno poznavanje koncepata računarskih sistema
- Prethodno iskustvo sa GPU programiranjem ili veštačkom inteligencijom nije potrebno

3. Tehnologije

C/C++, HIP, PCI, ROCm, LLVM, Triton, GDB i alati za merenje, profilisanje i praćenje izvršavanja GPU programa.

4. Teme kursa

- Uvod u savremene primene grafičkih procesora, od obrade slike do izvršavanje modela veštačke inteligencije.
- Osnovni elementi GPU programskog modela potrebni za razumevanje izvršavanja programa.
- Jednostavni HIP programi kao polazna tačka za proučavanje GPU hardvera i sistemskog softvera.
- GPU kerneli, niti, blokovi i organizacija paralelnog izvršavanja.
- Veza između programskog modela i GPU hardvera.
- Grupisanje niti u wavefront-ove i raspoređivanje posla na izvršne jedinice.
- Osnovni elementi GPU arhitekture i način izvršavanja instrukcija.
- Memorijska hijerarhija GPU-a: registri, lokalna memorija, keš memorije i globalna memorija.
- Upravljanje memorijom i prenos podataka između CPU-a i GPU-a.
- Osnovna uloga PCIe interfejsa i komunikacija GPU-a sa ostatkom računarskog sistema.
- Put programa od izvornog koda do izvršavanja na GPU hardveru.
- Uloga programskih prevodilaca u pripremi programa za izvršavanje na GPU-u.
- Uloga runtime sistema, drajvera i specijalizovanih biblioteka.
- Praktično posmatranje i istraživanje ponašanja GPU sistema kroz kratke programske primere, zadatke i eksperimente.
- Osnovni alati za merenje, profilisanje i praćenje izvršavanja GPU programa.
- Sistemska podrška za računске operacije koje se pojavljuju u neuronskim mrežama i drugim aplikacijama veštačke inteligencije

5. Literatura

1. *Accelerated Computing with HIP*, odabrana poglavlja.
2. Zvanična dokumentacija za HIP i ROCm: <https://github.com/rocm/>

-
-
3. Dodatni materijali, primeri i uputstva biće dostavljani studentima tokom kursa.