

Bevezetés a Komondor szuperszámítógép GPU partíciójának felépítésébe és használatába

A Komondor szuperszámítógép Magyarország legnagyobb teljesítményű számítógépe, feladata a hazai kutatás-fejlesztés, innováció támogatása. HPE Cray által gyártott rendszer a top500.org szuperszámítógép listán jelenleg a 300. helyet foglalja el több mint 6 petaFlop/s számítási összteljesítményével. Architektúrája megegyezik az első exaFlop/s teljesítményű Frontier és több más, a TOP500 lista első 20 helyén található szuperszámítógépekével (LUMI, Perlmutter, Adatastra). Teljesítményének döntő többségét (4,6 petaFlop/s) a GPU partíció, illetve az abban található 232 NVIDIA A100 GPU gyorsítókártya adja. Ennek a bevezető workshopnak a célja a GPU partíció felépítésének, hardver-szoftver sajátosságainak és hatékony használatának ismertetése annak érdekében, hogy a jelenlegi és jövőbeni felhasználók minél jobban ki tudják használni e rendszer képességeit feladataik megoldása során.

A képzés első modulja a GPU partíció hardver felépítését tekinti át. Szóba kerül a rendszer fizikai felépítése, a kabinet-blade-node hierarchia, a számítási node-ok felépítése. Kitérünk a rendszer CPU és GPU processzorainak részletes felépítésére, működési sajátosságaira is. Bár már egy GPU kártya is nagy teljesítményt nyújt, a szuperszámítógép lényege több GPU kártya együttes felhasználása nagy számítási feladatok megoldására. Emiatt részletesen tárgyaljuk a node-on belül található 4 db A100-as GPU kártya rendszer architektúráját, a CPU, illetve a GPU-k közötti összeköttetés rendszer (PCI-e, NVLink) működését és jellemzőit. Nagyobb feladatok esetén több számítási node egyidejű felhasználását, az azok közötti kommunikációt, a nagysebességű Slingshot 11 összeköttetés rendszer biztosítja. Mivel ez teszi lehetővé a szuperszámítógép-szintű feladatok végrehajtását, részletesen áttekintjük az összeköttetés rendszer architektúráját, funkcionalitását és egyéb fontos kommunikációs jellemzőit.

A második modul a GPU partíció szoftver környezetét tekinti át. Áttekintjük a GPU partíció használatához szükséges modul környezetet, a szükséges és opcionális szoftver modulokat, illetve a környezet konfigurálását. Kitérünk a Cray MPICH illetve OpenMPI szoftver könyvtárak összehasonlítására és használatára is, majd áttekintjük az NVIDIA HPC SDK legfontosabb építőelemeit. Megvizsgáljuk a multi-GPU programok lehetséges implementációs lehetőségeit, az ilyen programok fejlesztésének menetét, valamint a hardver és a szoftver szoros kapcsolatának következményeit. Az áttekintés egyik fontos célja, hogy azok a felhasználók is, akik maguk nem fognak ilyen programokat fejleszteni, csak kész, GPU támogatással rendelkező szoftvercsomagokat fognak használni, megértsék a rendszer alapvető működési elveit és tisztában legyenek azzal, hogy a GPU partíción futtatott programok milyen környezetben is futnak, attól mit várhatnak bizonyos típusú alkalmazások esetén.

A képzés harmadik modulja a kész programok futtatásával kapcsolatos lépések részleteit ismerteti. Áttekintjük a SLURM feladat ütemező használatát egy vagy több GPU kártyát használó programok futtatása esetén, a lehetséges konfigurálási opciókat, valamint a feladat végrehajtásával kapcsolatos fontos technikai részleteket (erőforrások allokálása a feladathoz, a CPU magok, GPU kártyák használata, esetleges megosztása vagy exkluzív használata, stb.).

Megvizsgáljuk az ideálistól eltérő párhuzamos futás okait, a várható számítási idővesztéseket okozó tényezőket. Felvázoljuk a programok futási teljesítményének vizsgálati lehetőségeit, majd az erre épülő program-optimalizálási lépéseket.

Célunk, hogy a workshop után a fejlesztők és végfelhasználók egyaránt pontosan értsék a rendszer működését, a felépítéséből-architektúrájából adódó következményeket mind a programok végrehajtása, de tervezése és implementálása szempontjából is; valamint, hogy a megszerzett ismeretek birtokában majd hatékonyabban tudják használni a Komondor GPU partícióját.