

Farkas Attila

Tudományos segédmunkatárs
HUN-REN SZTAKI

Budapest, 2024.11.28

Tanulságok és jövőkép

Tanulságok

HUN-REN Cloud jövőkép

- **Az utóbbi félévben nagyban növekedett a GPU erőforrások iránti igény**
 - Ezt a jelenleg érvényben lévő projekt és erőforrás allokálási rendszerrel nem lehet kiszolgálni
- **2025 január 01-től új erőforrás allokációs rendszert vezetünk**
 - Részleteket a HUN-REN Cloud weboldalán fogunk közölni és erről tájékoztatást is küldünk a felhasználóinknak
- **PaaS jellegű GPU hozzáférés**
 - A jelenlegi IaaS GPU hozzáférés kiegészítése
 - SLURM feladat ütemezőn keresztül
 - HPC kompatibilitás (pl.: Komondor) biztosítása, könnyebb átjárhatóság
- **Több szintű erőforrás igénylét tervezünk**
 - Pl.: 1 hónapos, 3 hónapos, 1 éves
 - Köteget jellegű hozzáférés
- **GPU erőforrások növelése**
 - Pályázati forrásokból
- **Célunk egy dinamikusabb GPU allokálási rendszer kialakítása**
 - A tényleges GPU kihasználtság emelése mellett

- **GINOP Plusz pályázati forrásokból**
 - Várhatóan 2025 tavaszi indulással
- **GPU erőforrások bővítése**
 - Több GPU kártya
 - Több VRAM GPU kártyánként
 - Multi-GPU támogatása virtuális gép szinten
- **Több memóriát tartalmazó virtuális gépek támogatása**
 - Felhasználói igények alapján
- **Quantum erőforrások elérésének támogatása**
 - Referencia architektúra alapon
 - Központilag menedzselt quantum erőforrás hozzáféréssel

AI4Science program támogatása

- Központi AI szakértők támogatása GPU erőforrásokkal
- GenAI4Science szolgáltatás biztosítása
 - <https://science-cloud.hu/genai4science>
 - Nyílt forráskódú nagy nyelvi modellek elérése



Köszönöm a figyelmet!

Kérdések és válaszok

HUN-REN
Magyar Kutatási Hálózat

<https://hun-ren.hu>