

Ütemezők a HUN-REN Cloud-on

A klasszikus HPC-től a konvergens infrastruktúrákig

Mező György, HUN-REN CSFK, HUN-REN WIGNER 2025.11.13.

Történeti áttekintés – A kezdetek

A mainframe korszak és a batch feldolgozás:

- IBM System/360, Cray-1: Lyukkártyás örökség
- Operátori döntés helyett szoftveres vezérlés: Queue (várakozási sor)
- Korai szabvány: NQS (Network Queueing System)

Az MPP (Massively Parallel Processing) kihívása:

- Több ezer "hétköznapi" processzor együttműködése
- Az egyszerű szekvenciális batch már nem elég
- Megoldás: PBS (Portable Batch System) – NASA Ames fejlesztés

Az ütemezők evolúciója

A PBS ökoszisztéma:

- TORQUE (nyílt forráskód) + Moab/Maui („agy”)
- Kereskedelmi vonal: Altair PBS Professional

Versenytársak:

- LSF (Load Sharing Facility) - ma IBM Spectrum LSF
- SGE (Sun Grid Engine): Fejlődése az Oracle felvásárlás után megtört

Az ütemezők evolúciója

A modern kor két főszereplője:

- Slurm: Az SGE utáni űrben emelkedett fel (SchedMD). Ma a TOP500 szuperszámítógépek több mint felén fut.
- HTCondor: 1980-as évek, Wisconsin–Madison Egyetem (2025-ben 40 éves). High-Throughput Computing (HTC) fókusz, heterogén erőforrások, „matchmaking”.

Ütemezők megjelenése a HUN-REN Cloud-ban (1.)

Igény a virtuális klaszterekre az önálló VM-ek mellett

WIGNER (Vukics András):

- Igény: Slurm alapú klaszter párhuzamos számításokhoz
- Megoldás: IaC (Infrastructure as Code) fejlesztés – kódból induló Slurm klaszter
- Eredmény: Emődi Márk közreműködésével a mai referencia-architektúra alapja

Agyi Képalkotó Központ (Keresztes Attila):

- Igény: fMRI képek feldolgozása (FSL szoftvercsomag)
- Megoldás: HTCondor alapú klaszter grafikus felülettel (sok független job futtatása)

Ütemezők megjelenése a HUN-REN Cloud-ban (2.)

Archeogenomikai Intézet (Szécsényi-Nagy Anna, Gerber Dániel):

- Igény: Nagy mennyiségű DNS-szekvenálási adat feldolgozása
- Megoldás: Bioinformatikai pipeline-hoz illeszkedő HTCondor klaszter

Csillagászati Intézet (Marco Pignatari, NuGrid):

- Igény: Hordozhatóság (Portability). Kódfejlesztés felhőben -> futtatás szuperszámítógépen (pl. Komondor)
- Cél: Felhős és HPC Slurm környezet szinkronizálása

Ütemezők megjelenése a HUN-REN Cloud-ban (3.)

Új GPU projekt igénylési rendszer HUN-REN CLOUD:

- Igény: Legyen a CLOUD felhasználóknak könnyen hozzáférhető GPU futtatásra alkalmas klaszter
- Megoldás: GPU-s VM-ek SLURM klaszterben konténerekre felkészítve és monitorozással

A Bridge Projekt – Konceptió

A két felhőág (WIGNER és SZTAKI) és a HPC közötti átjárhatóság

Két párhuzamos klaszter típus:

- HTCondor klaszter: High Throughput (HTC). Egymástól független feladatok, ideális a felhőmodellhez
- Slurm klaszter: HPC feladatok. Szorosan csatolt, elosztott memória-paradigmájú programok (MPI)

Összekötő elem: HTCondor-CE (Compute Element):

- Belépési pont
- A „router”, ami a job-okat irányítja.
- Összeköti a két világot.

A Bridge Projekt – Működési modell

Feladatok szétválogatása (Példa: fMRI feldolgozás):

- Előkészítés: Több alany képeinek párhuzamos előkészítése HTCondor
- Elemzés: Diffúziós képek elemzése (összetett számítás) -> Slurm

Architektúra:

```
User Job → HTCondor-CE → Decision Engine → Route to:
```

```
└─ HTCondor Pool (High-throughput)  
└─ SLURM Cluster (HPC/Memory-intensive)
```

A Bridge Projekt – Előnyök és lehetőségek

Rugalmasság:

- Több Slurm klaszter is beköthető (pl. felhős GPU rész + szuperszámítógép GPU partíció)

Cloud Bursting:

- Felszabaduló erőforrások gyors bevonása (pl. saját laptop erőforrásainak 50%-a a klaszterhez adva)
- Spot instance-ok használata számítások felskálázására

Továbblépés feltétele:

- Közös platform kialakítása az üzemeltetőkkel
- Közös AAI (Authentication and Authorization) infrastruktúra

Számítási infrastruktúrák konvergenciája – Kiindulópont

Klasszikus HPC:

- Homogén csomópontok
- Alacsony késleltetésű hálózat (InfiniBand, Slingshot)
- Batch ütemezők (Slurm) és párhuzamos fájlrendszerek (Lustre, GPFS)

Felhő (Cloud):

- Rugalmas kiszolgálás, heterogén hardver
- Erős virtualizáció, önkiszolgáló portál
- Orkesztrátorok (Kubernetes) ütemezők helyett (Service vs. Job)

„Egymással ortogonális megközelítések”

A konvergencia motorja

Technológiai háttér:

- Moore-törvény lassulása / Dennard-skálázás vége
- Teljesítménynövekedés kulcsa: Célspecifikus gyorsítók (akcelerátorok) és energiahatékonyság

Kereskedelmi felhőszolgáltatók válasza (On-premise HPC integráció):

- Azure: InfiniBand hálózat HPC-hez, ... , full stack HPC
- GCP: Parallelstore (DAOS) tárolás, ... , full stack HPC
- AWS: AWS ParallelCluster (HPC orkesztrátor), ... , full stack HPC

Szoftveres ökoszisztémák a konvergenciában

Altair Ecosystem:

- AAI-tól a Cloud Bursting-ig
- Interaktív HPC és költségoptimalizált migráció
- PBS Professional: HTC feladatok integrálása

IBM Spectrum LSF Suite:

- Hasonló funkciók a HPC és Cloud világ összekötésére. Mindkét esetben teljes vertikális integráció

Európai példa: JUPITER (Jülich)

Moduláris Szuperszámítógép (AI Factory):

- Nem kereskedelmi szolgáltatói példa
- Nyitott új modulokra (Neuromorphic, Quantum)

Dual-stack filozófia:

- JUPITER szuperszámítógép rész: AI training fókusz
- JARVIS AI Inference Cloud: OpenStack + Kubernetes orkesztrátor
- A klasszikus HPC és a modern felhő technológiák szimbiózisa egy helyen

Összegzés – A jövő iránya

Composable Infrastructure (Összeállítható infrastruktúra):

- A feladatok dinamikusan oda kerülnek, ahol a leghatékonyabban futtathatók
- HTC és HPC feladatok transzparens elosztása
- Felhő és Szuperszámítógép határainak elmosódása

A HUN-REN Cloud szerepe:

- Híd a kutatói igények és a heterogén erőforrások között
- Modern ütemezési technológiák (Slurm, HTCondor, Kubernetes) integrátora