



ClusterGrid

Szalai Ferenc

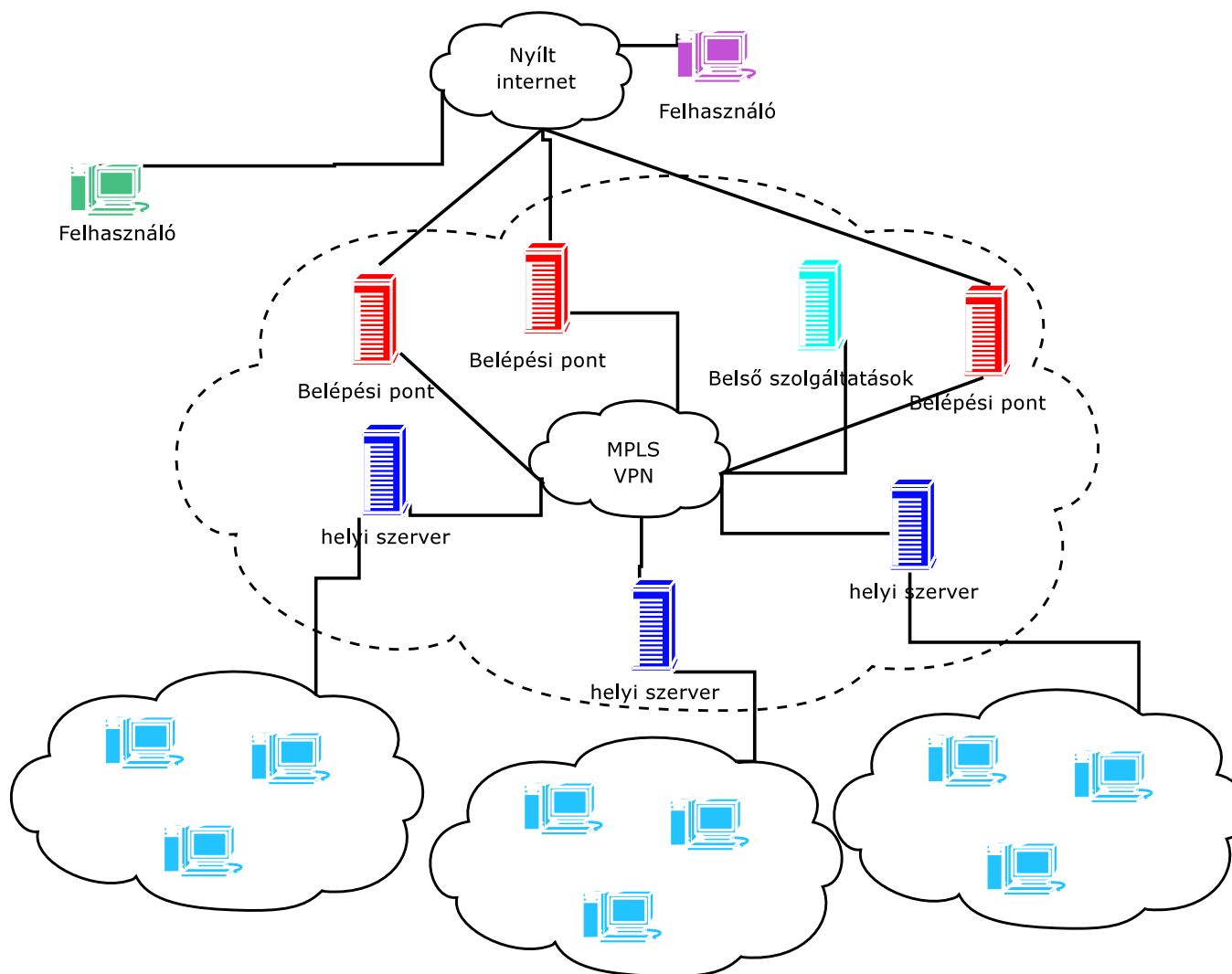
szferi@niif.hu

NIIF, MTA SZFKI

Bevezető

- 2002 július: OM beszerzési pályázat kb. 2000 gépre, 20 gép laboroként egyetemek, főiskolák, közgyűjtemények számára.
- Ötlet: Nappali üzemód mellett egy éjszakai (grid) is, a szabad kapacitások kihasználására.
- Feladat: Kidolgozni egy rendszert ami az ország teljes területén biztonságos és nagy megbízhatóságú szolgáltatást nyújt.

Hálózati Architektúra



Szoftver környezet

- Operációs rendszer: Linux, Debian disztribúció.
- Vékonykliens technológia alkalmazása, minden klaszter esetén.
- Feladat ütemező (bróker): NIIFI saját, nyílt forráskódú fejlesztése.
- Gyors alkalmazás és rendszerfejlesztés széles körben jól ismert technológiákra alapozva (apache, php, postgres SQL)
- Rendszermonitorizás
- Checkpoint technológia alkalmazása

ClusterGrid Bróker

- Mit akar a felhasználó? **Futtatni egyszerűen!**
- "Hagyományos mosópor" nem vált be: Globus, Condor stb.
- Nagy megbízhatóság, hibatűrés
- **Saját megoldás:** egyszerű és hatékony legyen
- Egyszerű felhasználói interfész:
 - Parancssori: clgr_info, clgr_status, clgr_submit, clgr_getout, clgr_rm
 - Web Portál.

Mi a feladat (job)?

- Nem csak egy futtatható állomány (exe)
- **Feladat** (job): szabványos könyvtárstruktúra + leíró állomány
- Lehet párhuzamos vagy szekvenciális

```
entry:~/cm_fm$ ls -l
bin/   bin/cm_fm
input/ input/test1.par input/test2.par
output/
pars/
submit
```

Job leíró állomány

```
entry:~/cm_fm$ cat submit
[run1]
executable = cm_fm
arguments = -c input/test1.conf
type = seq
[run2]
arguments = -c input/test2.conf
```

Komponensek

- **UI:** felhasználói programok
- **Submit:** interfész a belépési pontokon a felhasználói (interaktív belépes is). Grid szintű ütemezés, moduláris szerkezetű ütemezővel (core + tetszőleges nyelven írt ütemező modulok)
- **Exec:** Erőforrásokról és a futó feladatokról információt szolgáltat a Submit komponensek. Biztonságos futtató környezet.
- **Lokális ütemező interfész:** az Exec és a lokális ütemező közötti réteg.
- **Lokális ütemező:** klaszter szintű ütemezés, pl: Condor, SGE (nem a bróker része!)

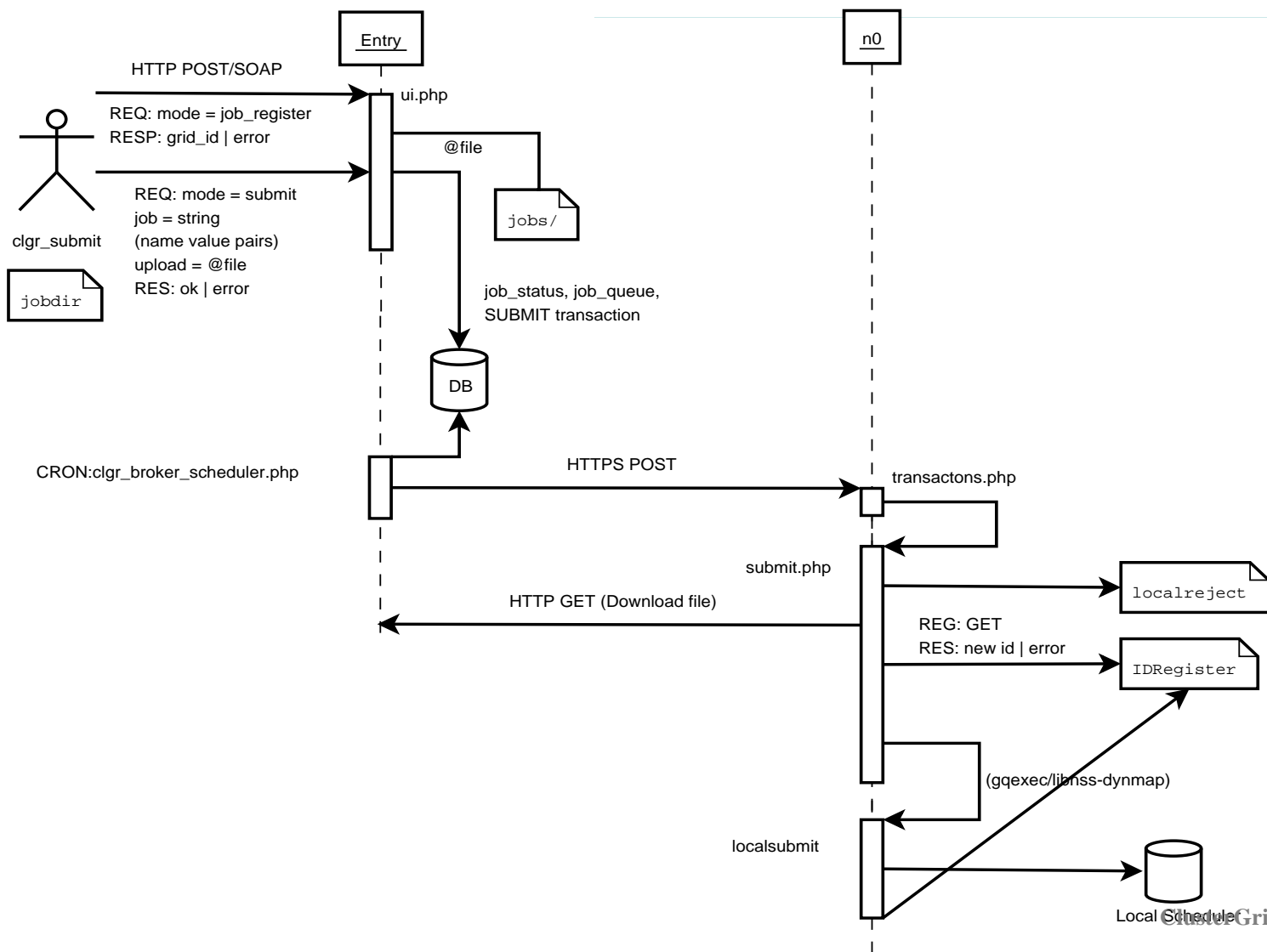
Alkalmazott technikák

- **Kommunikáció:** HTTPS POST
- **Tranzakciók:** Minden job-al végezhető feladathoz
 - indítás: submit
 - leállítás: remove
 - eredmény: results
 - állapot: status
- A tranzakciók, a feladatok és az erőforrások adatai SQL adatbázisban tárolódnak.

Biztonság

- **PKI:** Felhasználók, erőforrások azonosítása X509 cert.-el, saját CA (folyik az átálas egy EGrid PMA által elismert országos CA-ra)
- **IDRegister, NSS-Dynmap:**
 - erőforráson minden alkalmazás szeparált
 - grid felhasználók csak a belépési pontokon léteznek
 - Job alapú azonosítás
- Authorizációs technikák bevezetése és tervezése folyamatban

Példa



Eredmények

- Szolgáltatás szerű, működő, nem projekt grid rendszer kialakítása az akadémiai kutató közösség számára.
- kb 1000 csomópont, kb 500 GFlop számítási kapacitás.
- Több 100K feladat sikeres futtatása. Tudományos publikációnak alapja.
- Szunnyadó tudás felszínre kerülése.

Hova tovább hovatovább? - Bróker

- **Web services interfész:** együttműködés más rendszerekkel
- Monitoring rendszer integráció
- Különféle feladattípusok támogatása
- XML alapú feladat és erőforrás leírás és adatkezelés
- Biztonság növelése

Hova tovább hovatovább? - Rendszer

- Újabb több ezer csomópont bekötése.
- Szolgáltatás minőségének javítása, felhasználók oktatása.
- Új technológiák kiprobálása: CoLinux, OpenMosix, IPv6, stb.
- Széles nemzetközi együttműködés (EGEE, NoduGrid, stb)