



# Linux alapú Cluster rendszerek

Gáspár Attila (H3ABM8)  
Németh Lóránt (E9B0JL)

Budapest, 2003. január 25.

# 1. Az önálló laboratórium feladat témája

Három féléves munkánk során igyekeztünk megismerkedni a jelenleg széles körben elterjedt fűrtözési technológiákkal, illetve ezek közül is elsősorban a nagy rendelkezésre állású clusterekkel. Az egyes félévek tematikája a következőképpen alakult:

1. Convolo cluster rendszer folyamatainak modellezése
2. FailSafe és Convolo rendszerek összehasonlítása
3. Az „Informatikai infrastruktúra témalabor” tárgy Linux clusterek c. mérésének kidolgozása, feladatok összeállítása

Az első két félév munkájának jó összefoglalója a mellékelt mérési segédlet, amely olyan részletességgel taglalja a fenti témaköröket, hogy azok alapján a hallgatók képesek legyenek a technológia felhasználási területéről és jelentőségéről képet kapni, illetve a mérésen kiadott feladatokat önállóan megoldani.

Ezen összefoglalóban az utolsó féléves munka összefoglalásán túl szeretnénk a mérések jövőbeni vezetőinek segítséget nyújtani egy érdekes és színvonalas mérés összeállításához, illetve levezetéséhez. A „problémát” rögtön három részre osztanánk: a mérési környezet összeállítására, a mérési feladatok összeállítására, illetve a mérés levezetésére.

Mivel az elmúlt félévek során a FailSafe-et nem sikerült működőképes állapotba hozni ezért az alábbi fejezetek csak a Convolo installálására és konfigurálására vonatkoznak.

## 2. A mérési környezet összeállítása

### 2.1. Hardware beszerzés

Az általunk elvégzett és a későbbiekben ismerttetett mérési feladatok elvégzéséhez szükséges hardware-ek listáját az 1. táblázat mutatja. Természetesen amennyiben új feladatokat szeretnénk a mérésbe beilleszteni, vagy esetleg lehetőség adódik több rendszer párhuzamos összeállítására ezt a listát további elemekkel kell bővíteni, illetve egyes elemeket esetleg el lehet hagyni.

### 2.2. Hardware installálás

Mivel a felsorolt hardware-ek összerakása nem igényel semmiféle speciális tudást erre nem térünk ki részletesen. Az egyetlen dolog, amire figyelni kell,

2 db	IBM kompatibilis PC lehetőség szerint azonos HW és SW konfigurációval <ul style="list-style-type: none"> <li>• SCSI vezérlő</li> <li>• SCSI terminátor</li> <li>• 2db hálózati adapter<sup>1</sup></li> </ul>
1 db	„kereszt” UTP kábel <sup>1</sup>
2 db	„egyenes” UTP kábel
2 db	<i>szabványos</i> RS232 „nullmodem” kábel <sup>1</sup>
1 db	két utas külső SCSI diszk
2 db	power switch (WTI RPS-10 ajánlott)
1 db	PC hangkártyával
1-2 db	hangfal
1 db	stopperóra
1 db	projektor

1. táblázat. Szükséges hardware komponensek

az az, hogy a powerswitchek soros kábelei tényleg szabványosak legyenek, vagy a WTI honlapján leírtaknak megfeleljenek.

A powerswitchek hátulján található kis potméter szerű pöcköt legjobban, ha a 0 pozícióra állítjuk. Ez állítja be az eszköz azonosítóját. Elvileg mindegy mire állítjuk, de ugye Murphy-törvényei szerint...

### 2.3. Szoftver beszerzés

A Convolo jelenleg az összes<sup>2</sup> nagyobb Linux disztribúciót támogatja, azok legfrissebb verzióinak kivételével. Mindazonáltal amennyiben nincsen a hivatalosan is támogatott és az általunk választott újabb esetleg már telepített verzió között jelentős *glibc*-beli különbség, nyugodtan megpróbálhatjuk a telepítést a kiszemelt rendszerre.

Magából a Convolo rendszerből a Méréstechnika tanszék rendelkezik egy példánnyal, melyet a tanszéki adminisztráción lehet felvenni.

<sup>1</sup>Amennyiben soros heartbeat-et is szeretnénk elég lehet egy hálózati kártya, de ekkor kell még egy plusz „nullmodem” kábel és nem kell „kereszt” kábel

<sup>2</sup>RedHat, SuSe, Debain, Mandrake...

## 2.4. Szoftver installálás

1. Allítsuk be a BIOS-ban, hogy a gép a tápellátás visszatértekor auomatikusán elinduljon.
2. Installáljuk a kiválasztott operációs rendszert.
3. Ismertessük fel a SCSI winchestert (töltsük be a szükséges modulokat).
4. *fdisk* segítségével hozzuk létre az alábbi partíciókat:
  - I. 20Mb Linux native (*/dev/sda1*)
  - II. 20Mb Linux native (*/dev/sda2*)
  - III. 100Mb Linux native (*/dev/sda3*)
5. A */etc/init.d/rawio* fájlba írjuk bele a következő sorokat:

```
raw /dev/raw/raw1 /dev/sda1
raw /dev/raw/raw2 /dev/sda2
```

6. Gondoskodjunk róla, hogy a fenti fájl futtatható legyen és, hogy bootoláskor lefusson. Futtassuk le és a *raw -qa* paranccsal ellenőrizzük az eredményt.
7. A */etc/hosts* fájlba vegyünk fel egy *labpc1\_priv* és egy *labpc2\_priv* nevet két privát IP címmel. Ezek az IP címek és nevek fognak a két gépet közvetlenül összekötő hálózat interfészeihez tartozni.
8. A HA IP cím(ek)nek ha még nem lenne, kérjünk valamilyen néven egy bejegyzést a helyi DNS szerverre. (hamaster, haslave)
9. Helyezzük be a Convolo CD-t az egyik gép meghajtójába és indítsuk el a rajta található *install.exe* fájlt és értelem szerűen válaszoljunk a feltett kérdésekre.<sup>3</sup>
10. Végezzük el a fenti műveleteket a másik gépen is a partícionálás kivételével.
11. Futtassuk le az egyes számú gépen a */opt/cluster/bin/member\_config* parancsot és értelemszerűen válaszoljunk a feltett kérdésekre.

---

<sup>3</sup>amennyiben nem szeretnénk NFS-t mint HA szolgáltatást definiálni a clusterben, nyugodtan átléphetjük a kernel és kernelmodulok telepítését

12. A második gépen is futtassuk le a parancsot a `--init=/dev/raw1` opcióval és egyszerűen üssünk `entert` minden kérdésre.
13. A `/opt/cluster/bin/clu_check` parancs segítségével ellenőrizzük a konfigurációt.
14. A `/opt/cluster/bin/cluster -S` paranccsal ellenőrizzük a két powerswitchet.
15. Indítsuk el a cluster-t a `/etc/init.d/cluster start` paranccsal.
16. Amennyiben webszervert szeretnénk a clusterben használni:
  - (a) Telepítsük fel az apache-ot. Amennyiben PHP oldalakat is szeretnénk használni telepítsük fel a megfelelő modult is hozzá (`mod_php`).
  - (b) Hozzunk létre egy üres könyvtárat (pl.: `/var/www/cluster`).
  - (c) Az apache `DocumentRoot` változóját állítsuk ugyanerre a könyvtárra.
  - (d) Az eddig elvégzett műveleteket ismételjük meg a másik gépen is.
  - (e) mount-oljuk fel az adatpartíciót (`/dev/sda3`) a fenti könyvtárra.
  - (f) A kívánt fájlokat, weblapokat helyezzük el a könyvtárban.
17. Amennyiben MP3 lejátszással szeretnénk bemutatót tartani, úgy szükség lesz egy patche-lt xmms-re is, mely a forgalmazó cég honlapjáról [1] letölthető.

## 3. Mérés

### 3.1. Mérési feladatok

**1. feladat.** *A mintarendszer hardware-eszközeinek, illetve csatlakozásainak ismertetése*

- 2db PC
- 2db hálózati kártya gépenként
- 2db hálózati kábel, illetve 1db crosskábel
- külső SCSI diszk, kábelek, SCSI terminátorok
- powerswitchek, amik a másik gép soros portjára csatlakoznak

- powerswitchek sorba vannak kötve a gépekkel

**2. feladat.** *Egyszeres hibapont keresése*

Egyszeres hibapontok például:

- Közös fázisú tápellátás
- 1db fizikai diszk (nincs RAID)
- diszkeknek egy tápegysége van

**3. feladat.** *Lehetséges heartbeatvonalak felderítése a mintarendszeren*

- publikus hálózaton
- lokális hálózaton
- Quorum partíciókon

**4. feladat.** *Miben különbözik a laborban megépített rendszer a segédletben ismertetettől?*

Nincsen soros vonali heartbeat, de helyette van még egy hálózati heartbeatre lehetőség .

**5. feladat.** *Convolo felkonfigurálása member\_config segítségével*

**6. feladat.** *Grafikus management felület tanulmányozása*

**7. feladat.** *Apache szolgáltatás definiálása a grafikus felületen*

*Ha a szolgáltatás:*

- HTML oldal böngészése
- MP3 lejátszás

*Ha a failover-t előidéző ok:*

- Relocate
- SCSI kábel lehúzása <sup>4</sup>
- Gép kikapcsolása

**8. feladat.** *A MDT ingadozások okainak felderítése*

- Ha quorum partíció nem elérhető, de a többi heartbeat-vonal él, a Convolo másik gép lelövésének idejét kijebbb tolja.
- Timestamp-ek frissítési idejének gyakorisága.

---

<sup>4</sup>A kábel winchester felőli oldalát húzzuk le! A gép felőli oldal lehúzásának este **nem** tesztelt.

## 3.2. Beugró kérdések

1. Sorolj fel 3 különbséget a Convolo és a Failsafe között!
2. Mi az a Heartbeat funkció?
3. Milyen Heartbeat csatornák lehetségesek?
4. Milyen célt szolgál az osztott diszkrendszer a Convolo és a Failsafe esetében?
5. Mi a feladata a Quorum Daemon-nak?
6. Mi a feladata a Service Manager-nek?
7. Mi az a RawIO device, és a Convolo mire használja?

## 3.3. A mérés menete

A kiadható mérési feladatok mennyiségét és a mérés hosszát nagyban meghatározzák a rendelkezésre álló erőforrások. Amennyiben nem áll módunkban egynél több cluster-t összerakni úgy tűnik a legjobb megoldás, ha négy alkalommal egyenként négy darab négy fős csoport végzi a mérést. Ha még némi elméleti bevezetőt is szándékozunk tartani, úgy azt célszerű az egyes mérések elején a négy mérőcsoportnak közösen megtartani, csak úgy mint a beugrók megíratását. Ebben az esetben mindenképpen javasolt két előadó, és legalább 1 liter üdítő.

## 4. FailSafe

Az elmúlt félévben sajnos nem sikerült a FailSafe telepítésével előbbre jutnunk, a mérések előkészítése és levezetése miatt, így az alábbi információk a tavalyi tapasztalatokat igyekeznek összegezni.

Az első probléma amivel találgoztunk az az volt, hogy a FailSafe eleinte csak 2.2.3-as glibc-vel volt hajlandó futni. Negyedéves várakozás után megjelent a megfelelő patch, melyet a CVS-ből letöltött forrásra ráeresztve már működni kellett volna a rendszernek. Mielőtt tehát hozzáfognánk a telepítésnek nem árt ellenőrizni, hogy a rendszeren lévő glibc-vel hajlandó-e a FailSafe együttműködni.

Sajnos a szükséges csomagok listája nem teljes a dokumentációban. Amennyiben a dokumentációban még nem szerepel, más utalás, töltsük le és telepítsük a heartbeat-stonith csomagot, mivel e csomag nélkül a FailSafe nem fogja tudni kezelni a powerswitch-eket.

További probléma forrása lehet a grafikus kezelői felület, melyet nekünk a fél éve a laborban lévő gépeken nem sikerült beüzemelni, noha ugyanazt a Java futtató környezetet használtuk, mint az egyik kollégiumi gépen, ahol végül is sikerült elindítani. Ennek okát nem sikerült megtalálni, de esetleg érdemes lehet más ablakkezelőből is kiprobálni elindítani a GUI-t.

A *tcpmux* csomag fél évvel ezelőtti verziója nekünk hibásan működött és csak a forráskód módosításával sikerült rávennünk a működésre. Mielőtt a csomag feltelepítésének megörlődnénk nem árt tehát *telnet* segítségével bejelentkezni rá és pár egyszerű tesztet elvégezni.

Amit sikerült elérnünk:

- Szükséges csomagok lefordítása, telepítése
- Grafikus felület beüzemelése
- Cluster, erőforrások, erőforrás csoportok, erőforrás függések, szolgáltatás definiálása

Ami nem sikerült:

- A definiált szolgáltatás elindítása

Mivel ebbena félévben, mint azt már említettük nem tudtunk a FailSafe-el foglalkozni, nem tudjuk, hogy a legfrissebb verzió mennyire működőképes. Minden esetre a fejlesztői listán arra a kérdésre, hogy a fejlesztés jelenleg milyen tempóban folyik azt a választ adták az SGI munkatársai, hogy igazából nem fejlesztenek, mert a szoftver rendesen működik...

## 5. További tervek, lehetőségek

### 5.1. Shoutcast

Már elkezdtük tesztelni ezt a streaming média lejátszót. Ha sikerülne az apache szolgáltatáshoz valahogy hozzávenni, akkor kiküszöbölhető lenne a patch-el lejátszó. Sajnos jelenleg úgy tűnik csak Winamp-al lehet a streamet feltölteni rá. Ez még nem is lenne akkora probléma, de nem sikerült rájönnünk, hogy hálózati kiesés esetén, mi alapján dől el, hogy honnan folytatódik a lejátszás.

### 5.2. VMware ESX Server

Mivel a mérés egyik legnagyobb problémája az eszközök korlátozott száma, ennek a terveknek a segítségével lehetőség lenne egy fizikai gépen két virtuális futtatni. A termékről tudjuk, hogy lehetőséget nyújt virtuális hálózati



kártyák definiálására, melyen keresztül a heartbeat üzenetek mehetnek. Az egyetlen kérdés, hogy megoldható-e valamilyen módon a powerswitchek szimulálása...

### **5.3. Microsoft Clustering Technologies**

Szerettünk volna más platformokon elterjedt cluster alkalmazásokat is megvizsgálni, de sajnos erre sem volt időnk.

### **5.4. Load Balancing clusterek**

Elsősorban a Linux Virtual Server nevű alkalmazás keltette fel figyelmünket. Amellett, hogy ingyenes értesüléseink szerint hatékonyságban sem marad alul versenytársaival szemben. Mivel semmilyen speciális a laborban jelenleg nem elérhető eszközt nem igényel akár a mérés részét is képezhetné bemutatása.

### **5.5. HA Load Balancer**

A load balancing clusterek legfontosabb eleme a load balancer gép, aki a bejövő kéréseket osztja szét a fürt tagjainak. Érdekes kérdés lehet, hogy miként lenne növelhető a szolgáltatás rendelkezésre állása, ha egy másik géppel HA clustert alkotna.

# Tartalomjegyzék

<b>1. Az önálló laboratórium feladat témája</b>	<b>1</b>
<b>2. A mérési környezet összeállítása</b>	<b>1</b>
2.1. Hardware beszerzés . . . . .	1
2.2. Hardware installálás . . . . .	1
2.3. Szoftver beszerzés . . . . .	2
2.4. Szoftver installálás . . . . .	3
<b>3. Mérés</b>	<b>4</b>
3.1. Mérési feladatok . . . . .	4
3.2. Beugró kérdések . . . . .	6
3.3. A mérés menete . . . . .	6
<b>4. FailSafe</b>	<b>6</b>
<b>5. További tervek, lehetőségek</b>	<b>7</b>
5.1. Shoutcast . . . . .	7
5.2. VMware ESX Server . . . . .	7
5.3. Microsoft Clustering Technologies . . . . .	8
5.4. Load Balancing clusterek . . . . .	8

## Hivatkozások

- [1] [http://www.missioncriticallinux.com/products/mp3\\_failover\\_demo.php](http://www.missioncriticallinux.com/products/mp3_failover_demo.php)  
2003. január 26., 16:15