

Virtualizált szolgáltatási platform kialakítása Xen és AoE alapokon

Bevezető

- Mit?
 - intézményi vállalati szolgáltatásokat (web, állomány, levelező szerver, egyedi teszt rendszerek, virtuális oktatási hálózatok)
- Miért?
 - sok egyedileg felügyelt rendszer alacsony kihasználtsággal (< 10%)
 - sok vas, magas fenntartási költségek (pl.: áramszámla)
- Hogyan?
 - szabad szoftverekkel, nyílt technológiákkal

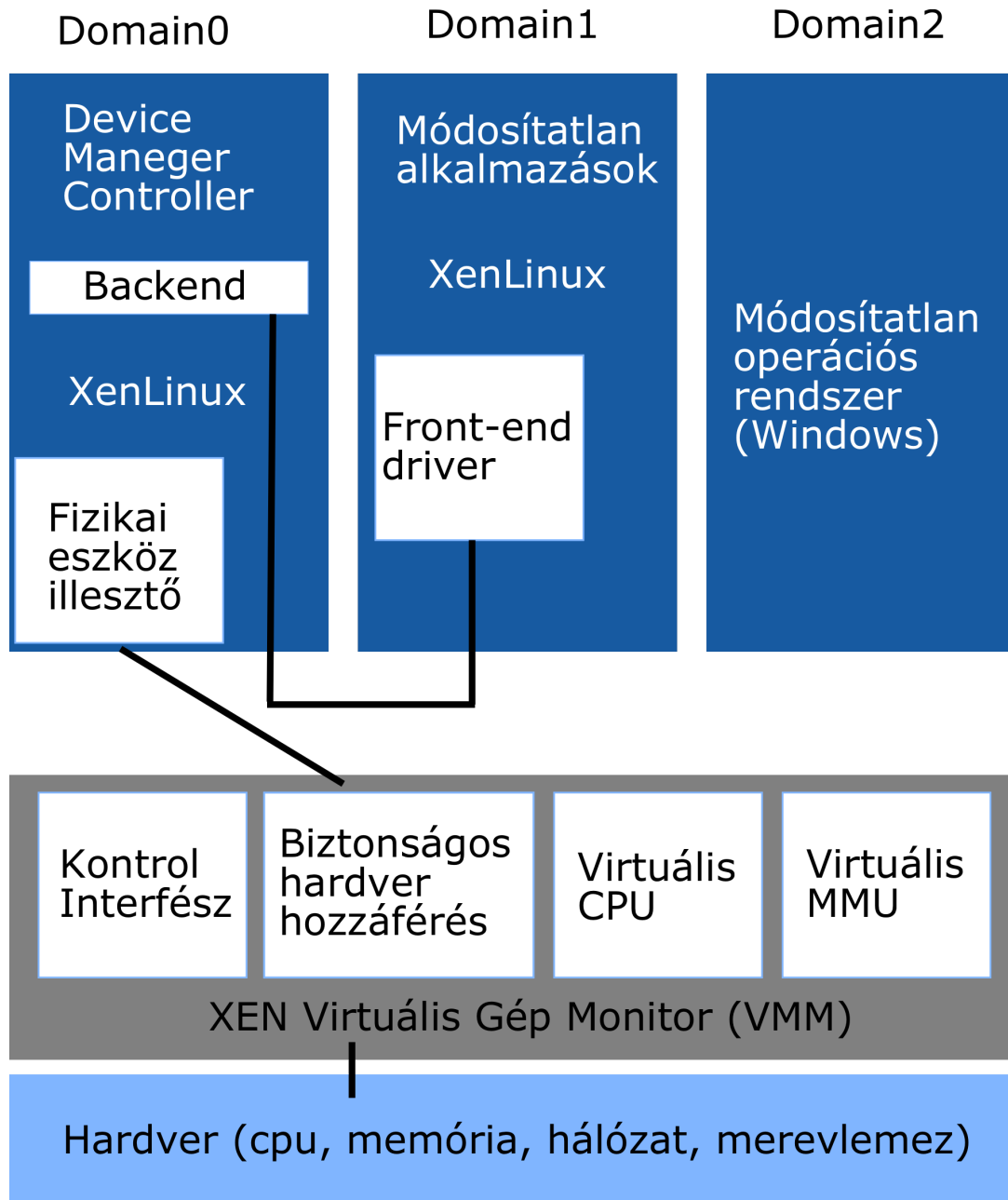
Virtualizáció - elvek

- **Virtuális gép:** absztrakció, a fizikai erőforrások megosztása, delegálása révén létrejött virtuális erőforrásokat használja.
- Virtualizációs technikák (x86):
 - alkalmazás szintű: pl.: cygwin, wine
 - operációs rendszer szintű: pl.: chroot, jail, openVZ, linux-vserver, solaris containers
 - emuláció: pl.: VMWare (sok okos, optimalizáció), QEMU
 - paravirtualizáció: pl.: **XEN**, User Mode Linux
 - hardverrel támogatott: Intel VT, AMD SVM

Virtualizáció - XEN

- **Hipervisor:** szuper-privilegizált módban futó kernel, közvetlenül a hardvert kezeli, erre portolják a virtualizált rendszereket
- Előnyök:
 - fizikai géphez közeli teljesítmény (2-3 % veszteség)
 - nyílt forrás, ipari támogatás (XenSource, IBM, SUN, Novell stb.)
 - linux terjesztések része (Debian, SuSE, RedHat EAL)
- Hátrányok:
 - operációs rendszert módosítani kell (no Windows)
 - nyílt management eszközök hiánya

Xen működése 1.



Xen kezelése (ízelítő)

```
# cat /etc/xen/noc.grid.conf
name = "noc.grid"
kernel = "/boot/vmlinuz-2.6.16-2-xen-vserver-686"
ramdisk = "/boot/initrd.img-2.6.16-2-xen-vserver-686"
memory = 256
vif = [ 'mac=00:16:3E:00:00:13, bridge=xenbr1' ]
disk = [ 'phy:/dev/xenimages/noc.grid,sda1,w' ]
root = "/dev/sda1 ro"
extra = "2"
on_poweroff = 'destroy'
on_reboot    = 'restart'
on_crash     = 'restart'
```

```
# xm list
```

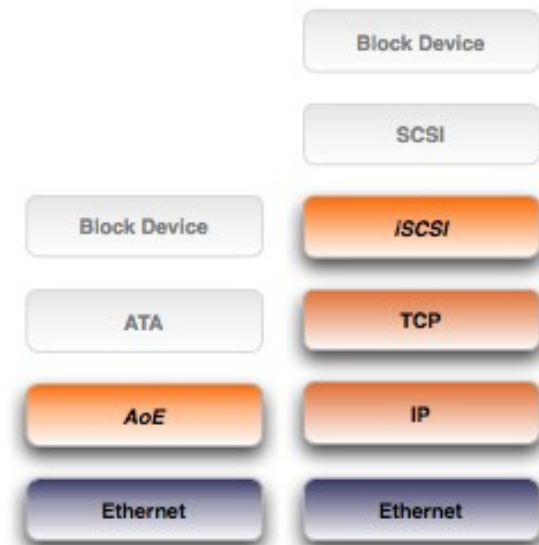
Name	ID	Mem(MiB)	VCPUs	State	Time(s)
Domain-0	0	250	4	r-----	47073.3
noc.grid	7	256	1	-b-----	101501.6
service.grid	8	128	1	-b-----	4728.6

Megfelelő tároló alrendszer 1.

- Hol tároljuk a virtuális gépek virtuális merevlemez tartalmát?
- Lehetőségek:
 - dom0-ba rakjunk sok lemezt: nehezen skálázható, nagy rendelkezésre állás kialakítása nehézkes
 - NAS (NFS): három egymástól független állományrendszer konzisztencia problémája
 - SAN: jó de általában drága (Fibre Channel)
 - IP/Ethernet SAN: perfect!

IP/Ethernet SAN

- forrás (target), cél (initiator) architektúra
- IP SAN: iSCSI
- Ethernet SAN: AOE (Ata-over-Ethernet)
- AOE: egyszerű protokoll ATA üzenetek natív ethernetén történő továbbítására
 - inititor 2.6.11 óta Linux kernel része
 - hardveres target (Coraid Inc.)



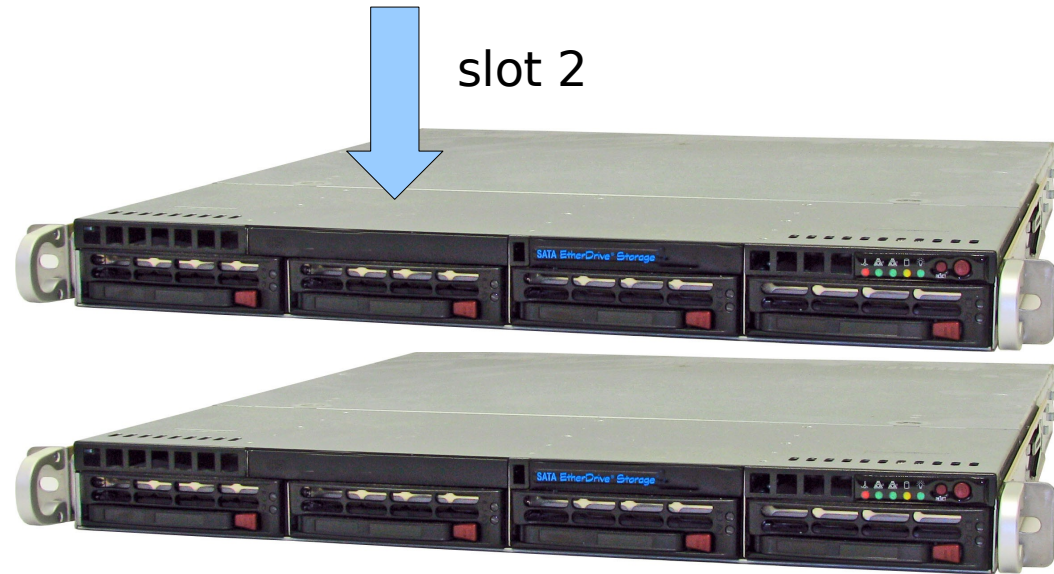
Munka AoE-val

- Szerver oldal:
- list
- make
- modprobe aoe
- aoe-stat

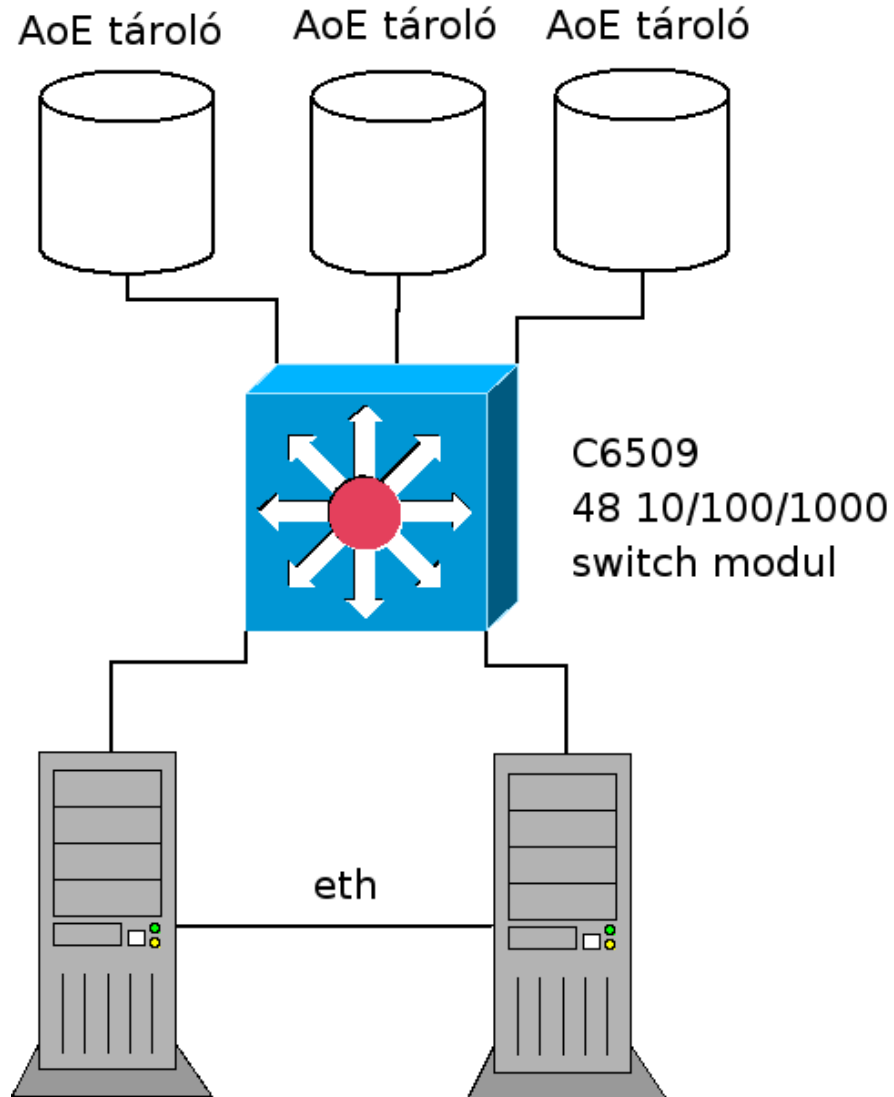
```
e1.0 500Gb eth1  
e1.1 500Gb eth1  
e2.1 1000Gb eth2
```

- mkfs.ex3 /dev/etherd/e1.2
- mount /dev/etherd/e1.2 /mnt

shelf 1



Nagy rendelkezésre állású virtuális kiszolgáló platform



Fizikai gép virtualizálása

- alkalmazások leállítása fizikai gépen
- dom0-án Volume kialakítása és csatolása (/mnt)

```
tar -czf - / | ssh dom0 'cd /mnt; tar -xzf -'
```

- hálózati interfészek beállítása dom0-ban (főleg, ha 802.1q VLAN technikát használjuk)
- xen config állomány elkészítése
- domain indítása (xm create domu.conf)

Összefoglalás

- Referencia:
 - NIF Intézet belső és külső szolgáltatási platformja 2 éve használja
- Tapasztalatok:
 - AoE: átlag 58 MB/s olvasás, 45 MB/s írás (eléri, néha meghaladja az EMC CX 500-at)
 - 99.98 % rendelkezésre állás (fizikai költözések és hardver hibák miatti kiesés)
 - 22 TB lemezterület hatékony felhasználása
 - sok küzdelem, de végül siker a valós idejű alkalmazásokkal (pl.: Asterix PBX)