

xen: serverkonsolidierung (nicht nur) für den haushalt

Linuxwochen Linz, 15.4.2005

christian mock
cm@quintessenz.org

inhalt

- was ist Xen?
- installation
- in der praxis
- was ich alles *nicht* ausprobiert hab

was ist Xen?

Xen is a virtual machine monitor for x86 that supports execution of multiple guest operating systems with unprecedented levels of performance and resource isolation. Xen is Open Source software, released under the terms of the GNU General Public License. We have a fully functional ports of Linux 2.4 and 2.6 running over Xen, and regularly use it for running demanding applications like MySQL, Apache and PostgreSQL. Any Linux distribution (RedHat, SuSE, Debian, Mandrake) should run unmodified over the ported OS.

<http://www.cl.cam.ac.uk/Research/SRG/netos/xen/>

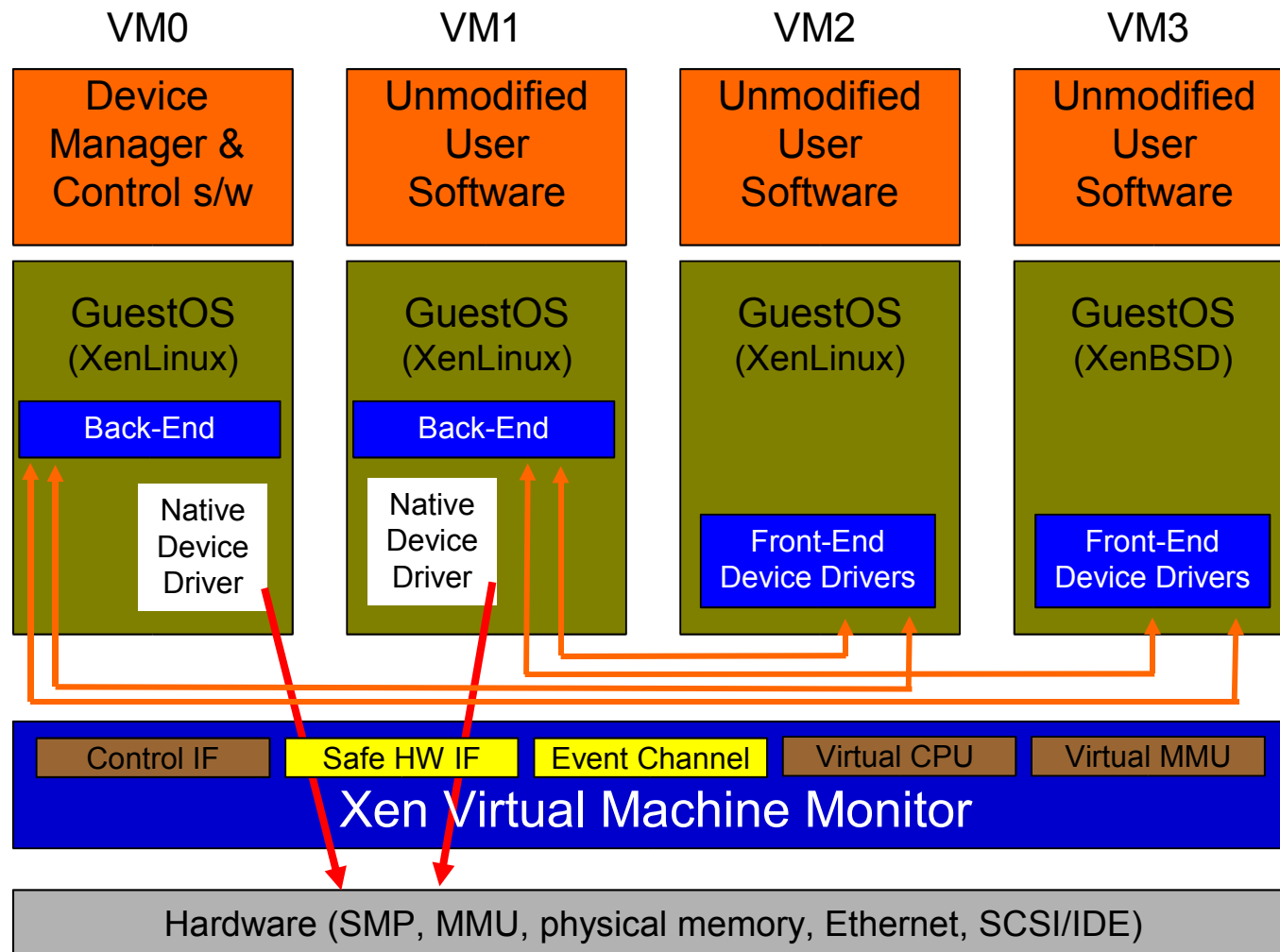
das erinnert mich jetzt an...

- VMware, User Mode Linux
richtig, Xen hat den selben zweck, erreicht ihn aber anders
- Bochs, qemu, Hercules
andere baustelle; das sind *emulatoren*, Xen virtualisiert nur die maschine

also wie jetzt?

- Xen ist ein virtual machine monitor
 - schlanker layer mit beschränkten funktionen
- guest-OS muß auf Xen portiert werden
 - wenig aufwand, memory mgmt, net + block drivers, ...
 - nur der kernel muß portiert werden
- verschiedene guest-OS parallel möglich
 - linux (2.4, 2.6), NetBSD, (FreeBSD, plan9)

ein bild sagt mehr...



(courtesy Ian Pratt, University of Cambridge and Founder of XenSource Inc.)

gut, aber wieso nicht (VMware|UML)?

- performance
 - I/O-performance (disk, net) ist bei Xen fast so gut wie native linux und drastisch besser als VMware und UML
- live-migration
 - virtuelle maschinen können (fast) ohne aussetzer von maschine zu maschine migriert werden

installation

- basis: Debian Sarge, Linux 2.6.10, Xen 2.0.4
 - get source
 - unpack
 - emacs xen-2.0/dist/install/boot/config-2.6.10-xen0
 - make world; make install
 - emacs /boot/grub/menu.lst
 - reboot

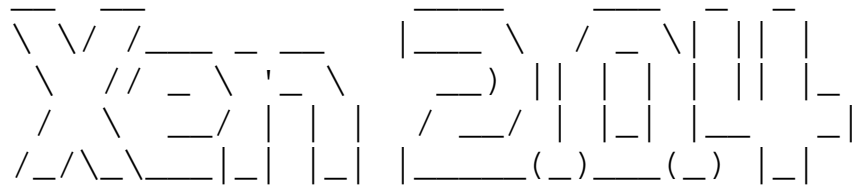
install: details

- sources: <http://xen.sf.net>
- linux-`{version}.tar.bz2` in xen-`{version}` directory kopieren
- defconfig für dom0 ist server-hardware-orientiert, anpassen
- installiert:
 - dom0, domU kernel
 - Xen VMM (`/boot/xen.gz`)
 - tools

bootbar machen

- Xen kann nur mit grub booten
- config laut doku
- `mv /lib/tls /lib/tls.disabled`
 - sonst performance schlecht
- reboot

my first Xen



<http://www.cl.cam.ac.uk/netos/xen>
 University of Cambridge Computer Laboratory

Xen version 2.0.4 (cm@tahina.priv.at) (gcc version 3.3.5 (Debian 1:3.3.5-8)) Wed Feb 16 21:33:54 CET 2005

Latest ChangeSet: information unavailable

(XEN) Physical RAM map:

```
(XEN) 0000000000000000 - 000000000009fc00 (usable)
(XEN) 000000000009fc00 - 00000000000a0000 (reserved)
(XEN) 00000000000f0000 - 0000000000100000 (reserved)
(XEN) 0000000000100000 - 000000002fffc000 (usable)
(XEN) 000000002fffc000 - 0000000030000000 (ACPI data)
(XEN) 00000000fec00000 - 00000000fec10000 (reserved)
(XEN) 00000000fee00000 - 00000000fee10000 (reserved)
(XEN) 00000000fff80000 - 0000000100000000 (reserved)
```

(XEN) System RAM: 767MB (786028kB)

(XEN) Xen heap: 10MB (10772kB)

(XEN) CPU0: Before vendor init, caps: 0183fbff 00000000 00000000, vendor = 0

(XEN) CPU caps: 0183fbff 00000000 00000000 00000000

(XEN) found SMP MP-table at 000f4ff0

und jetzt?

- Xen VMM hat gebootet
- linux dom0 kernel wurde geladen
- sollte normalen login-prompt zeigen
- momentan läuft nur dom0:

```
# xm list
Name                Id   Mem(MB)  CPU  State  Time(s)  Console
Domain-0            0     59      0  r----- 15796.7
```

- bingo!

domains anlegen

- was braucht man für eine domain?
 - disk
 - partition
 - loop-device
 - LVM logical volume
 - netzwerk
 - ip-adresse, MAC-adresse
 - config-file
 - /etc/xen/\${name}.xm

disk space

- partition (/dev/hdax)
 - sehr einfach
 - eher lästig (repartitionieren...)
- loop-device
 - leeres file anlegen
 - losetup /dev/loopx /path/to/file
 - lästig: numerierung
 - performance? (ein layer mehr)

disk space: LVM

- logical volumes anlegen
- $\{\text{domainname}\}-\{\text{funktion}\}$
 - z.B. news-swap, nms-root
- name ändert sich nicht
- performance ist gut
- platten dazu/weggeben...
- kein repartitionieren

disks: config

- im config-file:
 - `disk = ['phy:dataavg/news-root,sda1,rw']`
 - `root = „/dev/sda1 ro“`
 - `phy:blkdev-pfad,vm-name,flags`
 - mehrere möglich
 - root-angabe wie auf kernel-commandline
- domU greift dann über blockdev-treiber und dom0 auf disk zu

netzwerk

- domU eth0 ->
dom0 vifx.y -> bridge -> ethx -> kabel
- bridge muß nicht sein, routing geht auch
- setup durch scripts in /etc/xen/script

netzwerk: config

- config-file:

```
- vif = [ 'mac=de:ad:be:ef:37:16' ]
```

- kann wegfallen, dann MAC auto-generiert (bei jedem domU-boot)

- mehrere vifs möglich

```
- nics = n
```

- kernel-level DHCP (NFSroot) auch möglich

config-file

- examples:
`/etc/xen/{xmexample1,xmexample2}`
- kernel-image angeben
(`/boot/vmlinuz-n.n.n-domU`)
- memory-size festlegen
- name der domain

endlich starten...

```
# xm create -c test.xml
Using config file "test.xml".
Started domain test, console on port 9609
```

```
[...]
```

```
Kernel panic - not syncing: VFS: Unable to
mount root fs on unknown-block(8,1)
<0>Rebooting in 1 seconds..
^]
```

- ahja, kein filesystem...

wie OS installieren?

- installer, der nicht auf CD-ROM besteht
- disk in dom0 mounten, chroot+install
- dd von bestehendem FS
- Debian: debootstrap
 - debootstrap sarge /mountpoint
<http://ftp.at.debian.org/debian>
 - braucht nachbearbeitung
 - script: setup-xen-debian
<http://www.tahina.priv.at/~cm/talks/>

meine eindrücke

- stabil, performant
- angenehm: schnell neue maschinen aufsetzen
- funktionen damit leicht zu trennen
- spart eine menge hardware
- noch viel mehr möglichkeiten
 - zb jeder domU ein eigenes virtuelles LAN geben

was ich alles *nicht* ausprobiert hab

- migration
 - laufende guests von einer hardwaremaschine auf die andere verschieben, downtimes im ms-bereich
- driver domains
 - einzelne domU können hardware ansprechen und anderen zur verfügung stellen
- X11
- guest OS außer linux

roadmap für 3.0

- SMP guest support
- AMD64 support
- bessere tools (cluster mgmt)

schönen abend noch!

und danke für die aufmerksamkeit

folien:

<http://www.tahina.priv.at/~cm/talks/>