

# Implementace Xenu v univerzitním prostředí

Michal Švamberg

26. října 2007

- 1 Vlastnosti
  - Hledáme vhodné řešení
  - Výkonost
- 2 Xen na Západočeské univerzitě
  - Nastavení
  - Migrace
  - HW podpora virtualizace
- 3 Xen naživo

# Vlastnosti

Hledáme vhodné řešení

# Proč použít virtualizaci?

## Výhody:

- snadno dostupný stroj
- snapshoty
- duplikace
- šetří náklady a prostor
- lepší využití hardware
- a mnoho dalších

## Nevýhody:

- potřeba lépe dimenzovaného hardwaru
- vyšší závislost na spolehlivosti hardwaru

# Jakou virtualizaci použít (1)

## Emulace

- Bochs, QEMU, DOSEMU, PearPC
- plná virtualizace kompletního HW, dynamická rekompile
- není nutná úprava hostovaného OS, možný provoz jiného CPU
- velmi vysoká režie

## Nativní virtualizace

- VMware Server, Microsoft Virtual Server, HP Integrity Virtual Machine
- plná virtualizace s částečnou emulací HW
- nevyžaduje modifikace hostovaných OS, nutné zachovat CPU
- malá až střední režie

## Jakou virtualizaci použít (2)

### OS-level – virtualizace na úrovni operačního systému

- Linux-VServer, OpenVZ, Virtuozzo, FreeBSD Jail, FreeVPS
- virtualizace totožných OS na úrovni kernelu (hostujícího OS)
- nízká režie
- málo variabilní

### Paravirtualizace

- Xen
- neemuluje hardware, poskytuje speciální API
- vyžaduje modifikaci OS (neplatí v případě HW podpory)
- nízká režie (při užití HW podpory o něco vyšší)
- variabilnost použití, nutné zachovat CPU

## Jakou virtualizaci použít (3)

### Aplikační virtualizace

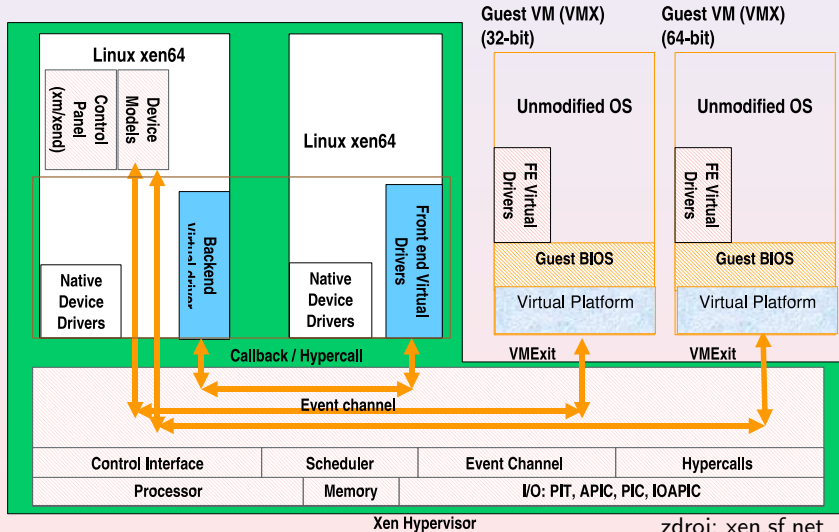
- Wine
- virtualizují se pouze základní komponenty nutné pro chod programu
- může být rychlejší běh než v originálním OS

## Proč jsme vybrali Xen

- podpora Linuxu
- nízká režie virtualizace
- snadný management (CLI)
- lze navázat na existující prostředí
  - souborový distribuovaný systém AFS
  - instalační metoda FAI
  - síťová infrastruktura používající VLAN (802.1q)
  - FibreChannel (FC) jako diskový subsystém
- cena



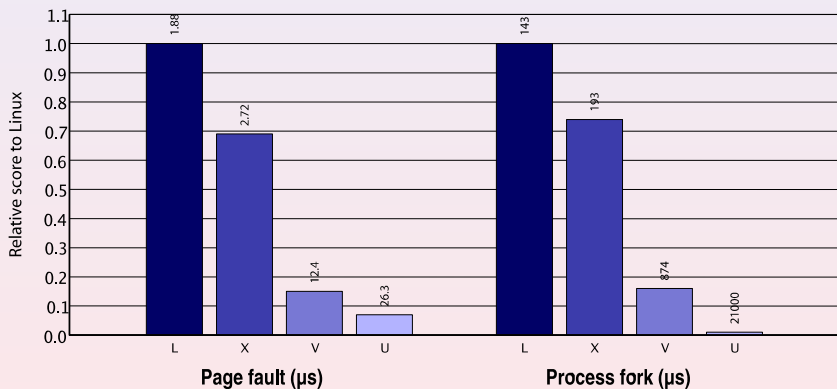
## Architektura Xenu



zdroj: xen.sf.net

# Výkonost

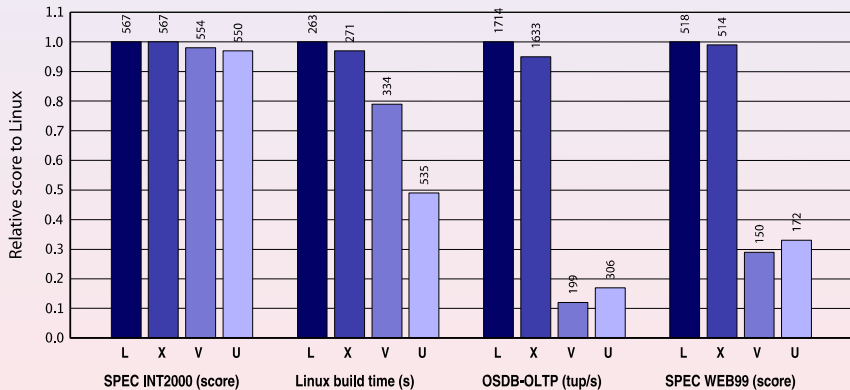
## Paměť



Imbench results on Linux (L), Xen (X), VMWare Workstation (V), and UML (U)

zdroj: xen.sf.net

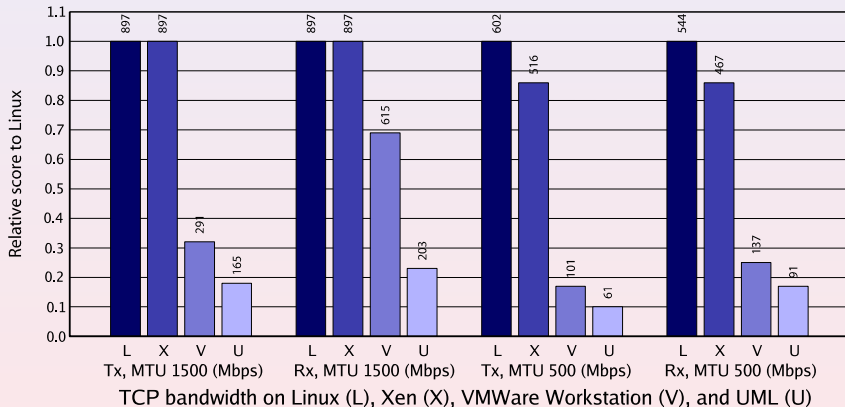
# Výpočetní testy



Benchmark suite running on Linux (L), Xen (X), VMware Workstation (V), and UML (U)

zdroj: [xen.sf.net](http://xen.sf.net)

# TCP testy



zdroj: xen.sf.net

# Xen na Západočeské univerzitě

## Nastavení

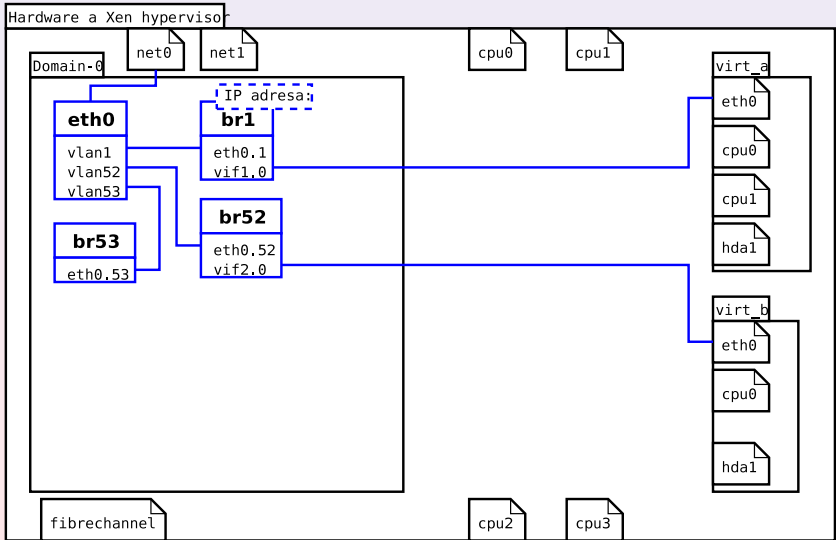
## Provoz virtuálních strojů

- 3x stroj hostující Xen (verze 3.0.4), 32bit
  - 2x DualCore Xeon na 3GHz
  - 8GB RAM, 2x Gbit ethernet, 2x 80GB SATA v SW RAIDu
  - FibreChannel karta
- 1x stroj hostující Xen (verze 3.1), 64bit, Vanderpool
  - 2x DualCore Xeon na 2,6GHz
  - 8GB RAM, 2x Gbit ethernet, 2x 160GB SATA v SW RAIDu
  - FibreChannel karta
- celkem 48 virtálních strojů (12 v ostrém provozu)
  - 64-850MB RAM
  - 10-30GB HDD (včetně 2GB swap souboru) na FC
  - 1x virtuální CPU
  - 1x virtuální ethernet připojen na bridge s danou VLANou
- správa virtuálních strojů
  - instalace FAI
  - konfigurace na AFS

# Nastavení sítě

- VLANy (správcovská je nativní jako záloha)
- bridge propojují 802.1q rozhraní s virtuálními interfaci (vif)





## Nastavení sítě – /etc/network/interfaces

```
auto br1  
auto br52  
auto br53
```

```
iface br1 inet static  
    address 147.228.1.25  
    netmask 255.255.255.0  
    broadcast 147.228.1.255  
    gateway 147.228.1.1  
    bridge_ports eth0.1
```

```
iface br52 inet loopback  
    bridge_ports eth0.52  
    up ip address del 127.0.0.1 dev br52
```

```
iface br53 inet loopback  
    bridge_ports eth0.53  
    up ip address del 127.0.0.1 dev br53
```

## Napojení na bridge – brctl show

bridge name	bridge id	STP enabled	interfaces
br101	8000.003048683896	no	eth0.101
br4	8000.003048683896	no	eth0.4
<b>br1</b>	8000.003048683896	no	<b>eth0.1</b> vif2.0 vif3.0 vif5.0 vif6.0 vif9.0
br52	8000.003048683896	no	eth0.52 vif7.0
br53	8000.003048683896	no	eth0.53
...	...	...	...

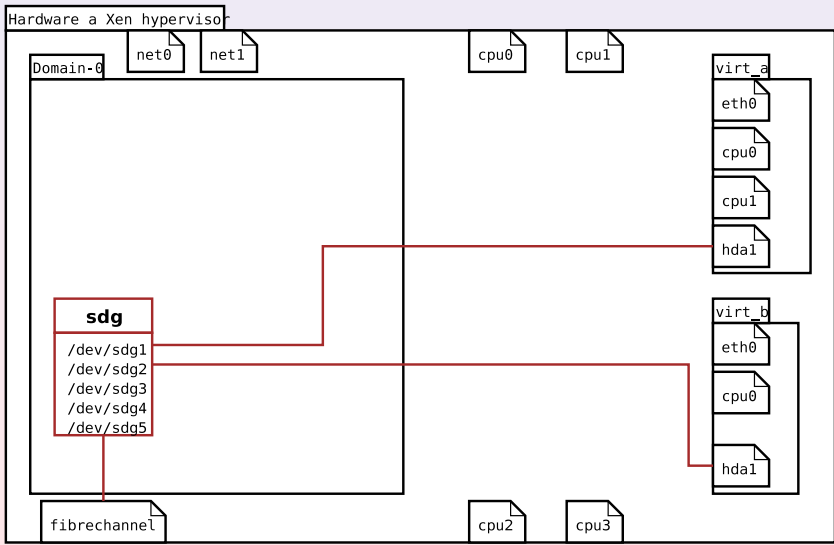
# Diskový subsystém

## Virtuální stroje

- virtuální stroj má vyhrazenou partitionu na FC (RAID 5)
- swap do souboru z důvodu migrace
- partition na FC 10, 20, 25 nebo 30GB
- celkem 4 disky z FC po 200GB (každá 15 partitions)

## Domain-0

- 2x80GB SATA disk
- všechny partitions včetně swapu v SW RAIDu (mdadm)
- 50+GB vyhrazeno jako lokální úložiště



# Processory

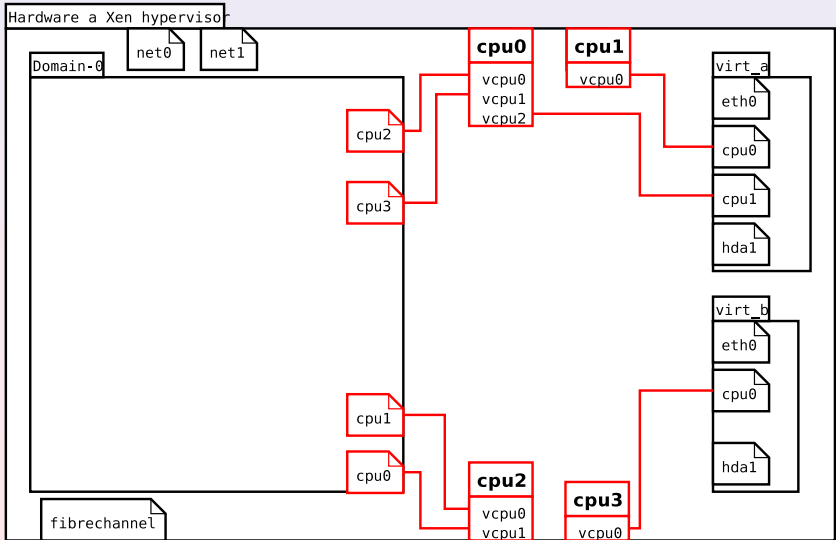
Virtuální stroj může mít libovolný počet virtuálních procesorů, každý z těchto virtuálních procesorů může být mapován na libovolný fyzický procesor. Platí také pro Domain-0.

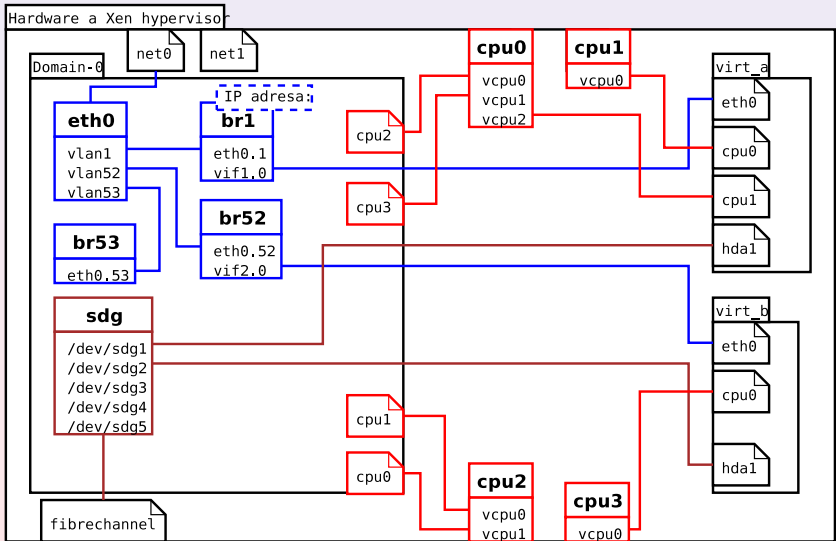
Současný stav:

- každý virtuální stroj s jedním virtuálním cpu

Výhled:

- několik vcpu v jednom virtálním stroji
- automatizované rozložení zátěže







# Konfigurace na AFS

AFS (Andrew File System) nám poskytuje několik výhod:

- je třeba releasovat při změnách
- RW volume pro editace, přípravu a testování
- RO repliky na 4 serverech v různých lokalitách
- jedna konfigurace dostupná na všech Domain-0

<http://www.openafs.org/>

# Instalace metodou FAI

FAI (Fully Automatic Installation) pro rychlou instalaci:

- instalační metoda pro Debian
- konfigurace uložena na AFS
- používáme vlastní mirror Debianu
- pro použití Xenu vlastní rozdělování disků – žádné

<http://www.informatik.uni-koeln.de/fai/>

# Konfigurace Xenu

## boot:

```
kernel = "/boot/xenu-linux-2.6.16.33-xenu"  
memory = 128  
name = "xenik"  
vif = [ 'mac=00:16:3e:05:30:71, bridge=br53' ]  
disk = [ 'phy:/dev/fc/fc-p1d1p13,hda1,w' ]  
root = "/dev/hda1 ro"
```

## install:

```
kernel = "/boot/xenu-linux-2.6.16.33-xenuboot"  
root = "/dev/nfs"  
extra = "ip=147.228.53.71:147.228.53.154:147.228.53.1:255.255.255.0:\  
xenik:eth0:off FAI_LOCATION=147.228.53.154:/usr/local/share/fai.dev \  
FAI_ACTION=install FLAGS=verbose,debug,createvt"
```

# Xen na ZČU

## Migrace

# Migrace

Migrace je přesun virtuálního stroje (hosta) mezi jednotlivými hardwarovými nody (hostitely).

Migrací virtuálních strojů získáme:

- vyšší dostupnost při údržbě
- možnost vyrovnávání zátěže nodů

Migraci rozdělujeme na:

- off-line: rychlejší, stroj je ale suspendovaný
- on-line (live): pomalejší, stroj je v provozu

# Prerekvizity pro migraci

- zachování síťového segmentu
- virtuální stroj umístěn na síťovém datovém prostoru:
  - NAS: NFS, CIFS
  - SAN: Fibre Channel
  - iSCSI, síťové blokové zařízení
  - drdb

## Jak migrace probíhá

Migrace z virtuálního stroje  $X$  z uzlu  $A$  na uzel  $B$ :

**pre-migration** aktivace  $X$  na uzlu  $A$ , výběr cíle na uzlu  $B$

**reservation** inicializace kontejneru pro  $X$  na uzlu  $B$

**pre-copy** cyklické kopírování "špinavých" paměťových stránek

**stop-and-copy** poslední fáze kopírování:

- pozastavení virtuálního stroje  $X$  na uzlu  $A$
- přesměrování síťového provozu
- synchronizace zůstávajícího stavu

**commitment** aktivace virtuálního stroje na uzlu  $B$ , uvolnění virtuálního stroje na uzlu  $A$

# Jak migrace probíhá

Migrace z virtuálního stroje  $X$  z uzlu  $A$  na uzel  $B$ :

**pre-migration** aktivace  $X$  na uzlu  $A$ , výběr cíle na uzlu  $B$

**reservation** inicializace kontejneru pro  $X$  na uzlu  $B$

**pre-copy** cyklické kopírování "špinavých" paměťových stránek

**stop-and-copy** poslední fáze kopírování:

- pozastavení virtuálního stroje  $X$  na uzlu  $A$
- přesměrování síťového provozu
- synchronizace zůstávajícího stavu

**commitment** aktivace virtuálního stroje na uzlu  $B$ , uvolnění virtuálního stroje na uzlu  $A$



# Z vlastní zahrádky

- Konfigurace

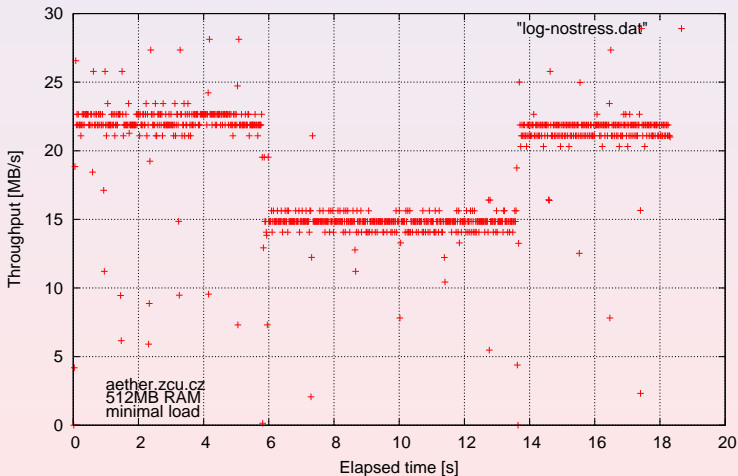
- Xen verze 2.0.7
- migrovaný stroj o velikosti 512MB (*aether.zcu.cz*)
- migrace probíhala ze stroje *xen2* na stroj *xen3*
- měřicí stroj (*phoebe.zcu.cz*) umístěn na *xen3*
- nutno nastavit forward delay: `brctl setfd br53 0`

- Měření

- z *aether* se tahá dokola jeden soubor
- z přírustku za 10ms se spočítá rychlost
- použít Perl s `Time::HiRes` a `LWP::Parallel::UserAgent`
- pro zatížení *aether* použít `stress --vm 3`

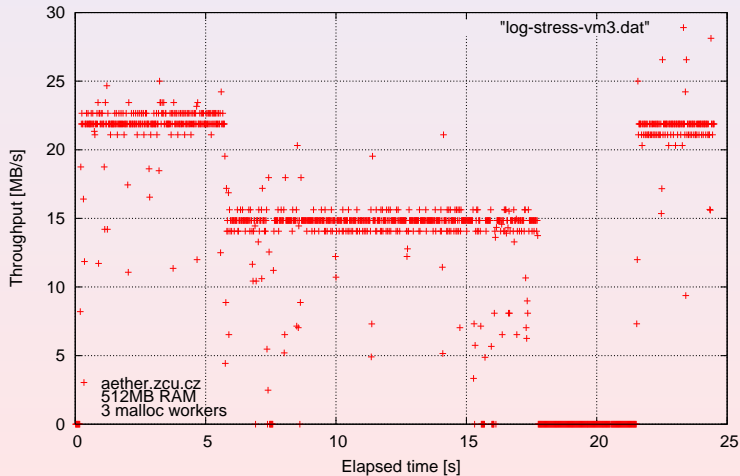
# Migrace nezatížené *aether*

Live migration - transmission rate progress  
*aether.zcu.cz* with minimal load



# Migrace *aether* se zátěží

Live migration - transmission rate progress  
aether.zcu.cz heavily loaded



# Xen na ZČU

## HW podpora virtualizace

# Podmínky

- podporu virtualizace v chipsetu a BIOSu na základní desce
- podporu virtualizace na CPU (vmx = Intel, vms = AMD)
- zapnutou HW podporu virtualizace v BIOSu
- verze Xenu 3.x
- zkompilevanou podporu HW virtualizace (xen-tools)

# Nastavení

```
kernel = "/usr/lib/xen-3.1-1/boot/hvmloader"  
builder='hvm'  
vif = [ 'type=ioemu, mac=00:16:3E:05:20:63, bridge=br52, \  
        model=ne2k_pci']  
disk = [ 'file:/mnt/data/pothos/pothos_C.img,ioemu:hda,w']  
device_model = '/usr/lib/xen-3.1-1/bin/qemu-dm'  
cdrom="/mnt/data/pothos/Windows.Vista.Enterprise.64-bit.CZ.iso"  
boot="cd"
```

# Ukázka teorie v praxi

Xen naživo

## Co uvidíme

- instalace virtuálního stroje
- migrace
- využití HW podpory virtualizace



## Použité odkazy



Miroslav Suchý: *Úvod do virtualizace pomocí XENu*

<http://www.root.cz/clanky/uvod-do-virtualizace-pomoci-xenu/>



Ian Pratt a kol.: *Live Migration of Virtual Machines*

<http://www.cl.cam.ac.uk/research/srg/netos/papers/2005-migration-nsdi-pre.pdf>



Kolektiv autorů: *Xen 3.0 User Manual*

<http://www.cl.cam.ac.uk/research/srg/netos/xen/readmes/user.pdf>



Vladimír Holub: *Virtualizace*

<http://www.roznovskastredni.cz/dwnl/pel2007/09/Holub.ppt>

Děkuji za pozornost.