

cin.ufpe.br



I/O Bound Performance Tuning for Cloud Experiments

Erico Augusto C. Guedes



UNIVERSIDADE FEDERAL DE PERNAMBUCO

0. Main Objective

I/O bound

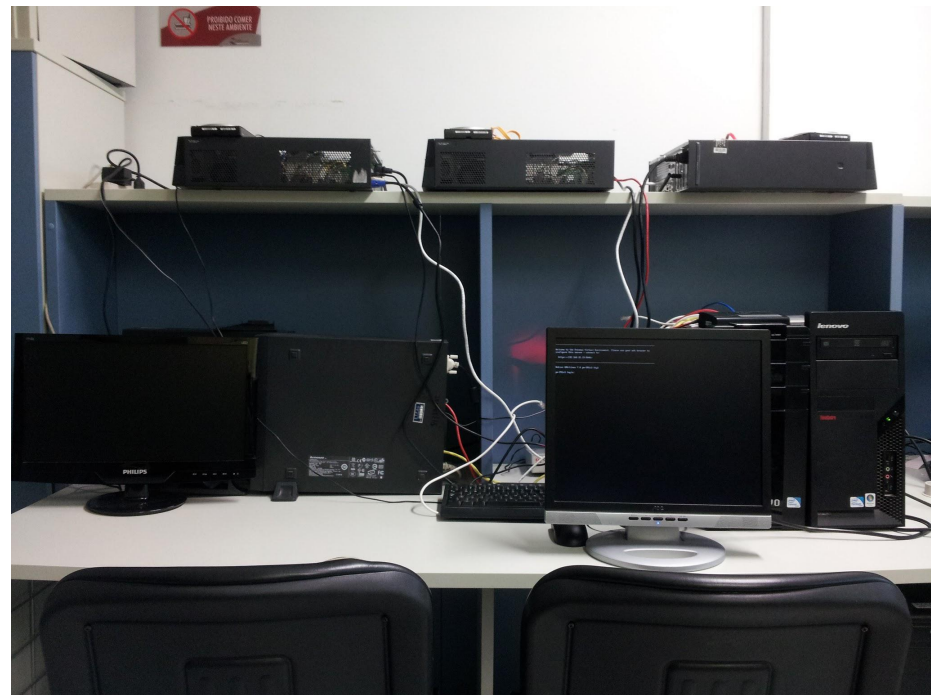
- **Is it possible to execute a virtualized service with similar performance of non-virtualized environment?**

1. Problem

- How to optimize I/O performance for experiments on virtualized environments?
 - Hardware
 - Software
 - Operating System
 - Hypervisor
 - Service

Hardware

- Any similarity?



Software

- **Operating System**

- **Processes priority**

- `nice`, `renice`

- **Filesystem > ext3**

- **Consistency mechanism: JOURNAL**

- `jbd2`

- **CAUTION: Overhead**

- **`noatime`, `nodiratime`**

- avoid to **update inode access** times on these file system, aimed to faster access.

- **ext4:**

- `extends`, `dealloc`, `mballo`

- improvements that reduces latency

Software

- Operating System:
 - SMP affinity of IRQs:

```
# cat /proc/interrupts

          CPU0           CPU1           CPU2           CPU3
0:         124             0             3             285   IO-APIC-edge     timer
1:           0             0             0             2   IO-APIC-edge     i8042
3:           0             0             0             6   IO-APIC-edge     tpm0
7:           1             0             0             0   IO-APIC-edge     parport0
8:           0             0             0            142   IO-APIC-edge     rtc0
9:           0             0             0             0   IO-APIC-fasteoi  acpi
12:          0             0             0             4   IO-APIC-edge     i8042
14:          0             0             0             0   IO-APIC-edge     pata_atiixp
15:          0             0             0             0   IO-APIC-edge     pata_atiixp
16:          0             0             2            425   IO-APIC-fasteoi  ohci_hcd:usb3,
ohci_hcd:usb4, snd_hda_intel
17:           0             0             0             0   IO-APIC-fasteoi  ehci_hcd:usb1
18:          0             0             2            1520   IO-APIC-fasteoi  ohci_hcd:usb5,
ohci_hcd:usb6, radeon
19:           0             0             0             2   IO-APIC-fasteoi  ehci_hcd:usb2
22:          0             198           79965          7619797   IO-APIC-fasteoi  ahci, ath
25:        5157           23328        18052003        319643548   PCI-MSI-edge     sky2@pci:0000:02:00.0
NMI:         164           136            99            375   Non-maskable interrupts
LOC: 60397496       50480973       36142997       87630731   Local timer interrupts
SPU:           0             0             0             0   Spurious interrupts
PMI:          164           136            99            375   Performance monitoring interrupts
IWI:           0             0             0             0   IRQ work interrupts
RES: 48541572       40170756       27700714       10065919   Rescheduling interrupts
```

Software

- Hypervisor
 - Full-virtualization



- Container-based OS virtualization(COS)

Software

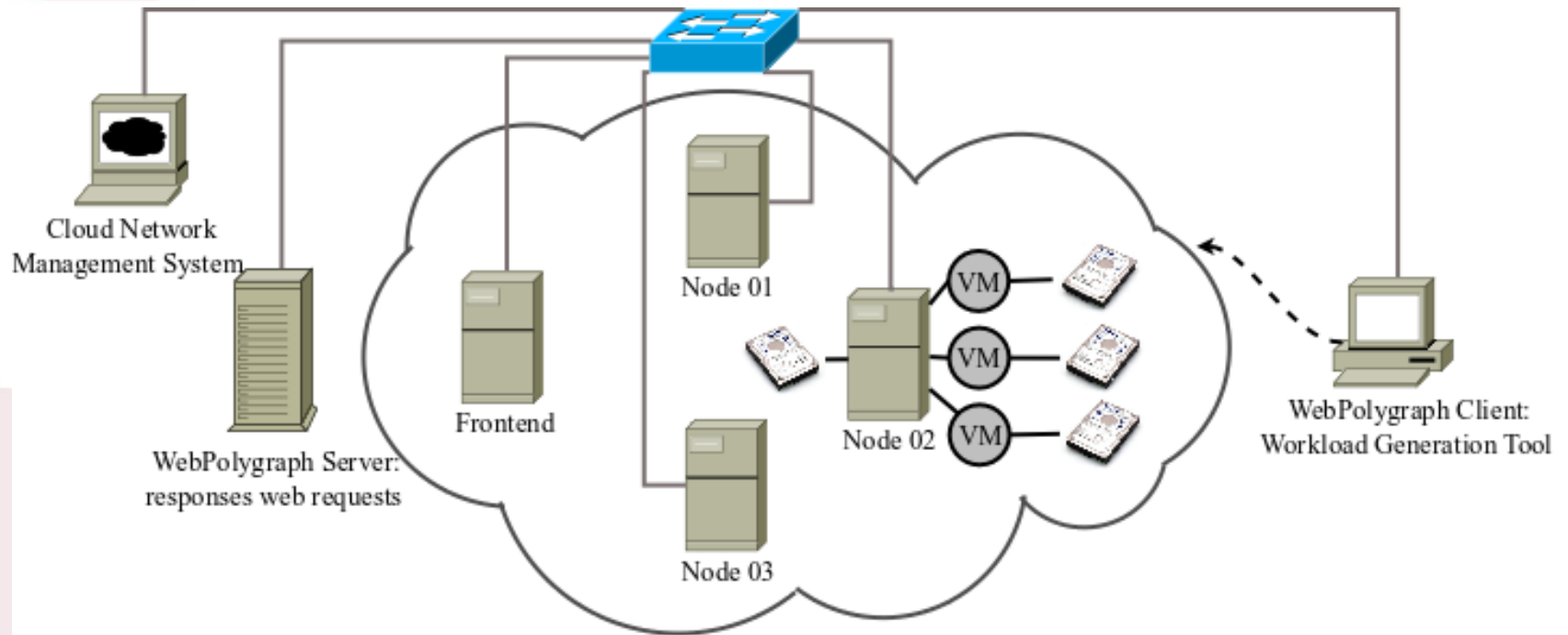


- **Service**

- Optimize service parameters
- Generic Example for I/O:
 - Tight partition and filesystem: fit partition and filesystem sizes to your needs.

2. Testbed

Service: Web cache proxying



3 Physical Machines X 3 VMs

DAS - Directed-Attached Storage

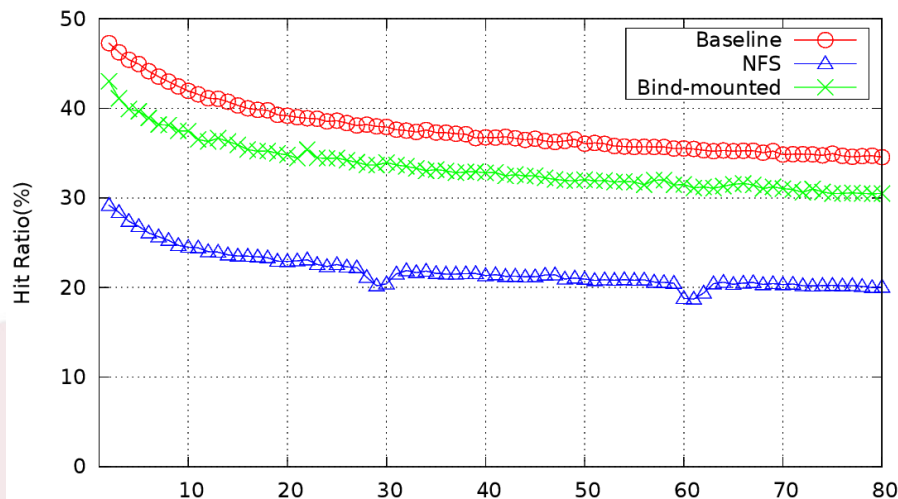
3. Results of optimizations



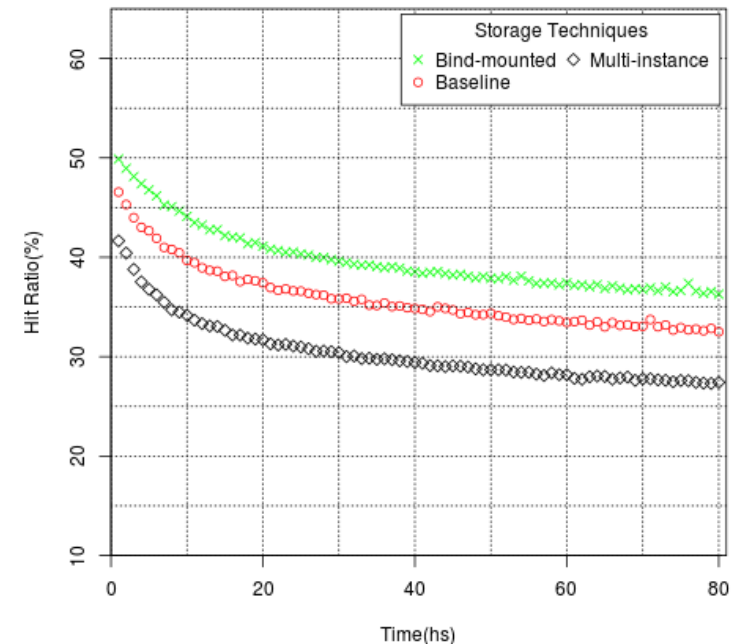
- **Start of experiments:**
 - I/O overhead of:
 - full virtualization : journal: lack of SMP affinity
 - **A = 58.676%**
 - **Better availability reached: 86.35%**
 - Forbids performance analysis
- After apply:
 - COS : journal disabling : SMP affinity
 - **A = 99.999% (Five 9's)**

3. Results of optimizations

- Performance Metric: Hit Ratio
 - Total of 3GB of cache: 1GB per node: IEEE ISCC 2014



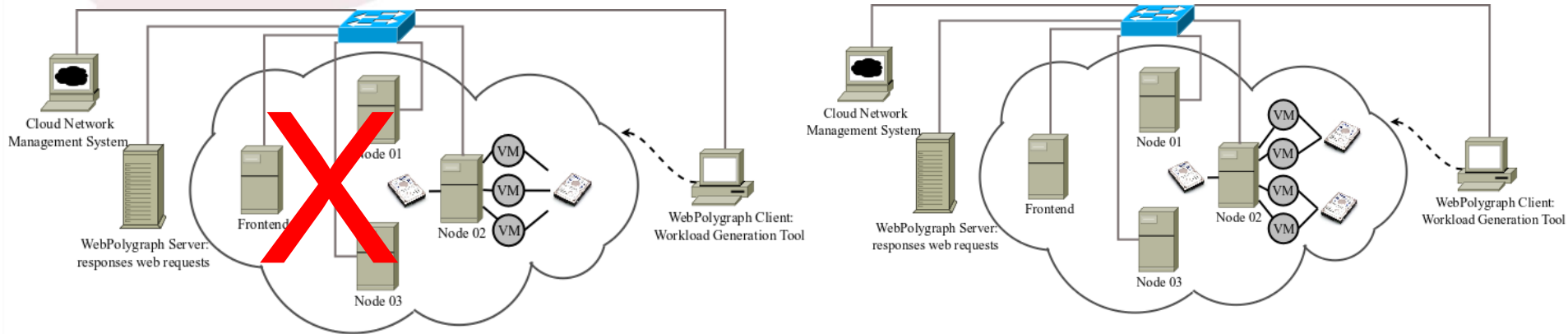
KVM -> OpenVZ; 1 disk per CT
(squid log disabled)



All set of performance optimization
(squid log disabled)

3. Results of optimizations

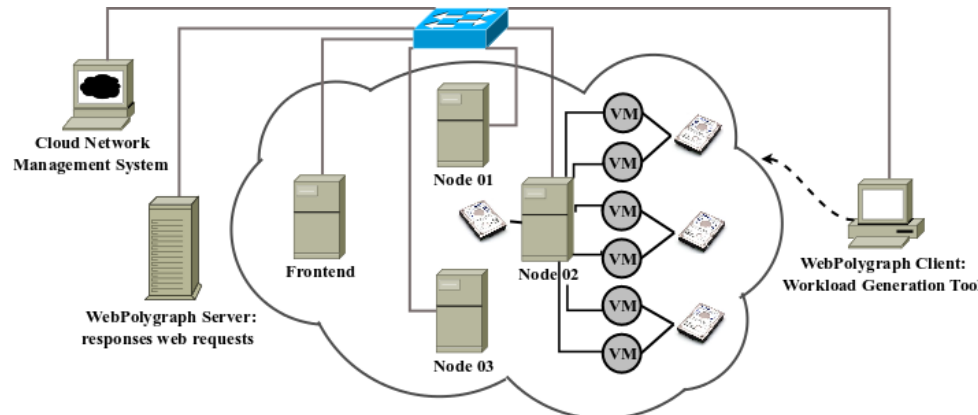
- Optimization: enabled scenarios



Cache of 300GB:
- 100GB per CT

Cache of 400GB:
- 100GB per CT

Squid log enabled: journal turned off



Cache of 600GB:
- 100GB per CT

I/O Bound Performance Tuning for Cloud Experiments

3. Results of optimizations



- Proxmox-OpenVZ: 18CTs: 1.8T of Total Cache
 - Capacity Planning

Browser address bar: <https://192.168.15.23:8006/#v1:0=-openvz%2F108:4::::11:>

PROXMOX Proxmox Virtual Environment
Version: 3.0-20/0428106c

You are logged in as 'root@pam' [Logout](#) [Create VM](#) [Create CT](#)

Server View: Container 108 ('CT108') on node 'pm-PM1v3'

Start Shutdown Stop Remove Migrate Console

Summary Resources Network DNS Options Task History UBC Backup Permissions

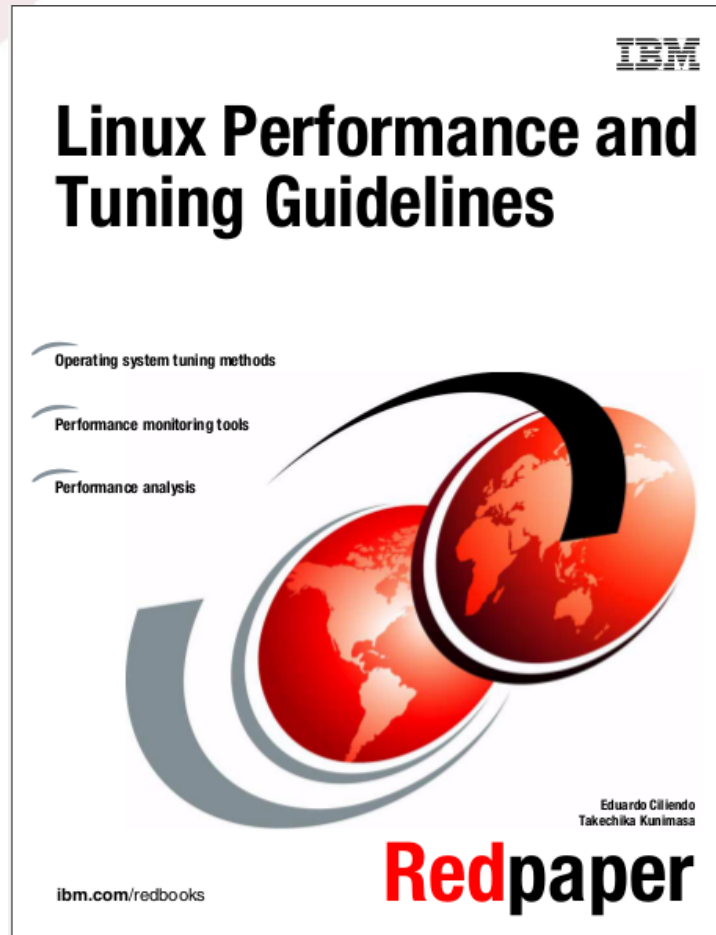
Add Remove Edit

Type	IP address/Name	Bridge	MAC	Host ifname	Host MAC
Network Device	eth0	vmbro	D6:EA:EC:61:8D:21	veth108.0	36:67:E1:50:23:0C

Tasks Cluster log

Start Time	End Time	Node	User name	Description	Status
------------	----------	------	-----------	-------------	--------

4. Optimization Reference



Questions/Suggestions

?



cin.ufpe.br



I/O Bound Performance Tuning for Cloud Experiments

Erico Augusto C. Guedes



UNIVERSIDADE FEDERAL DE PERNAMBUCO