Enhancing Cloud Efficiency

Proactive Memory Management

Jing Yan

Aug 10, 2024

Agenda

- Proactive memory management solution for cloud
- Demo: proactively moving inactive memory to different numa nodes

Proactive Memory Management

- Define your “Inactive Memory”
  - Age-based, Heat-class-based, etc.
- Identify your “Inactive Memory”
  - Accessed Bit in Page Table Entries, Active and Inactive Lists (eg, Least Recently Used algorithm), etc.
- Move your “Inactive Memory”
  - Swapping: swap inactive pages to swap space in disk, which will make room in physical memory for pages that are actively being used [1][2].
  - NUMA Optimization: move inactive pages to different numa nodes to help optimize memory access patterns, reduce latency and improve performance [1].

How Cloud Benefits from Proactive Memory Management?

Foundation of Cloud: Virtualization and Containerization

- Virtualization: memory overcommitment is very common for hypervisor.
- Containerization: containers use host's kernel.

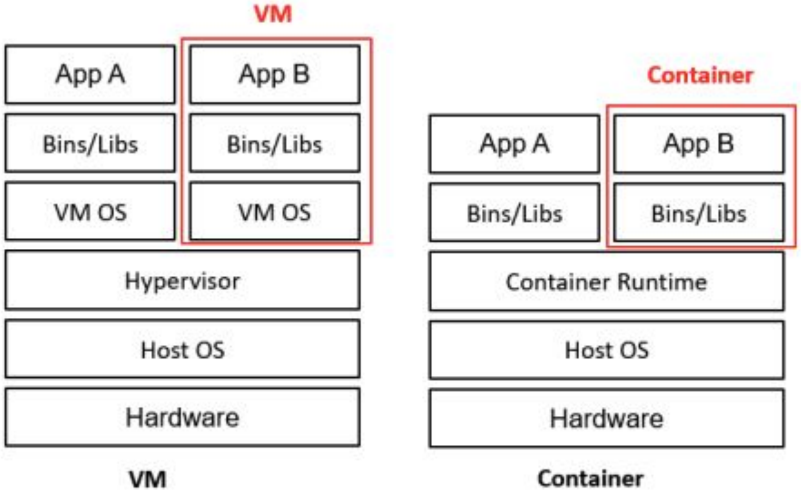


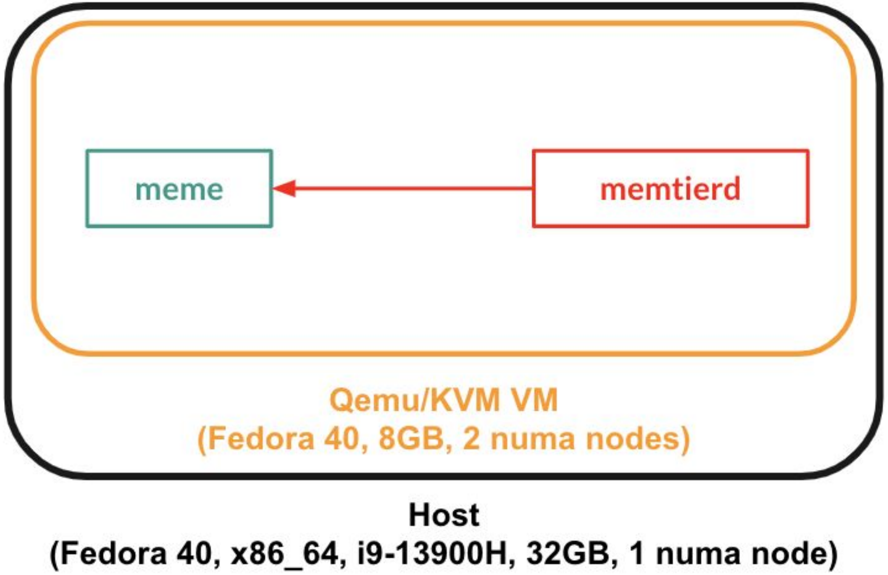
Figure: VM with Type-2 Hypervisor and Container [3]

Demo: proactively moving inactive memory to different numa nodes

Demo environment introduction

- **meme**: a process which allocates, reads, and writes memory

- **memtierd**: a damon process which monitors meme and move inactive pages of meme accordingly


meme
memtierd
Qemu/KVM VM
(Fedora 40, 8GB, 2 numa nodes)
Host
(Fedora 40, x86_64, i9-13900H, 32GB, 1 numa node)

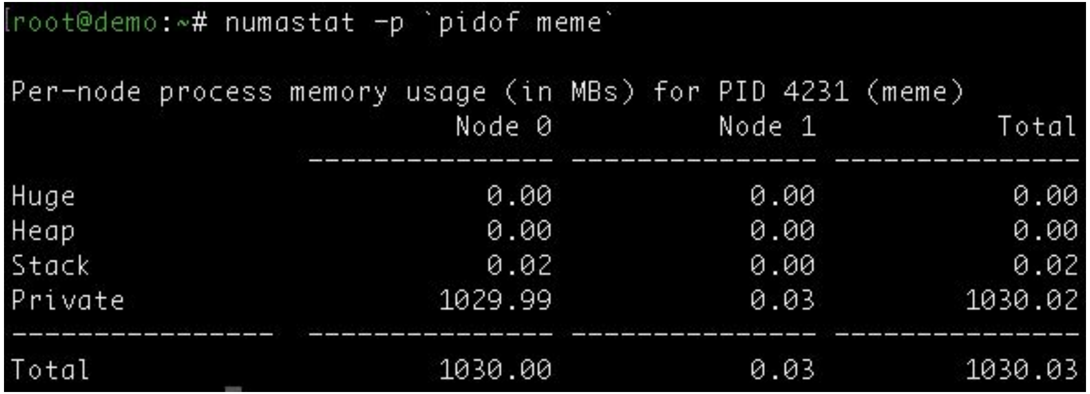
Demo (1)

# Create a new cgroup “meme”, and limit “meme” process which allocates 1GB and actively writes 300 MB to use “0” memory node only.

```
root@demo:~# mkdir -p /sys/fs/cgroup/meme
root@demo:~# echo 0 > /sys/fs/cgroup/meme/cpuset.mems
root@demo:~# meme -bs 1G -bwc 1 -bws 300M -ttl 2h &
echo `pidof meme` > /sys/fs/cgroup/meme/cgroup.procs
```

# Clean up

```
root@demo:~# killall meme
root@demo:~# rmdir /sys/fs/cgroup/meme/
```



```
root@demo:~# numastat -p `pidof meme`

Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                   1029.99            0.03         1030.02
----------------  --------------- --------------- ---------------
Total                     1030.00            0.03         1030.03
```

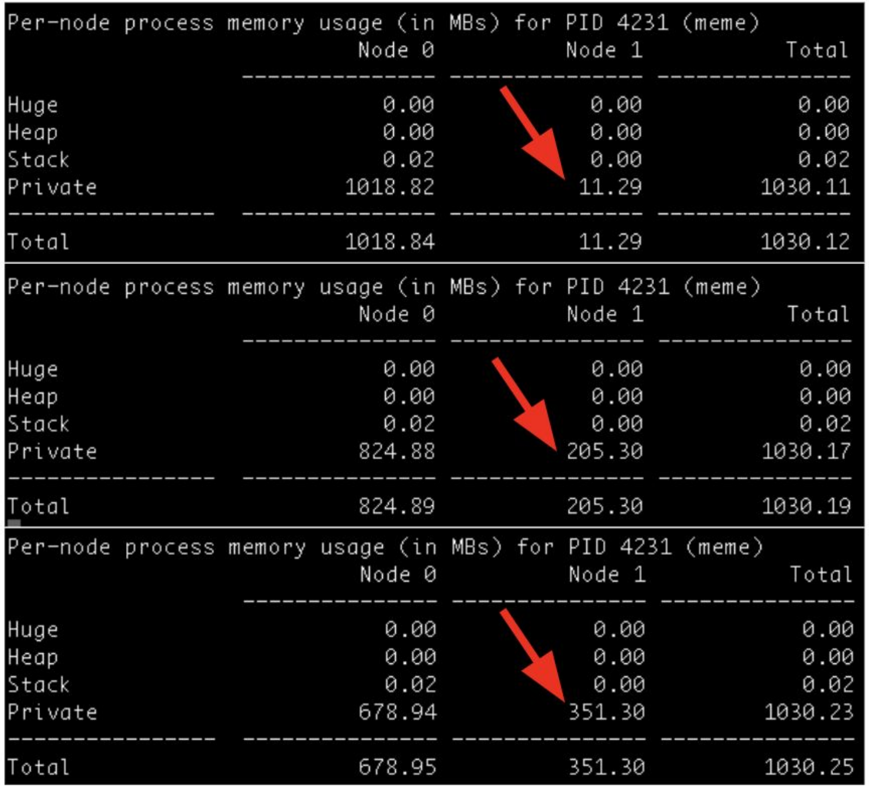
Demo (2)

# Enable reliable idlepage tracking and allow processe belongs to “meme” cgroup to use “0-1” memory nodes.

```
root@demo:~# echo 0 > /proc/sys/kernel/numa_balancing
root@demo:~# echo 0-1 > /sys/fs/cgroup/meme/cpuset.mems
```

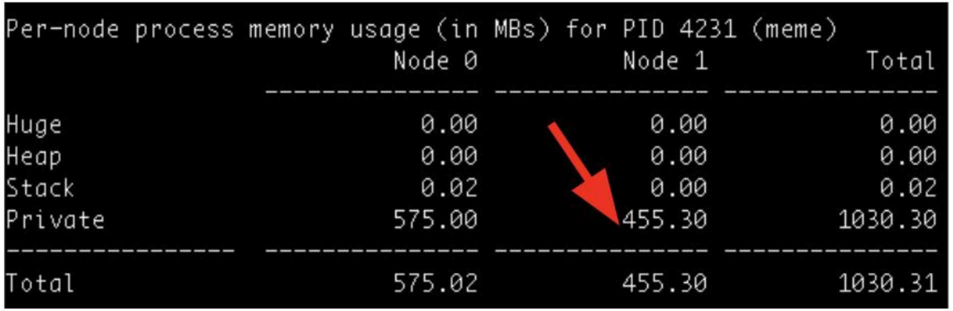
# Start “memtierd” to proactively moving inactive pages

```
root@demo:~# cat memtierd-age-idlepage.yaml
policy:
  name: age
  config: |
    intervalms: 5000
    pidwatcher:
      name: cgroups
      config: |
        cgroups:
          - /sys/fs/cgroup/meme
    idledurationms: 8000
    idlenumas: [1]
    tracker:
      name: idlepage   [Reference 4]
      config: |
        pagesinregion: 512
        maxcountperregion: 1
        scanintervalms: 4000
    mover:
      intervalms: 20
      bandwidth: 2000

root@demo:~# memtierd -config memtierd-age-idlepage.yaml -prompt
memtierd>
```

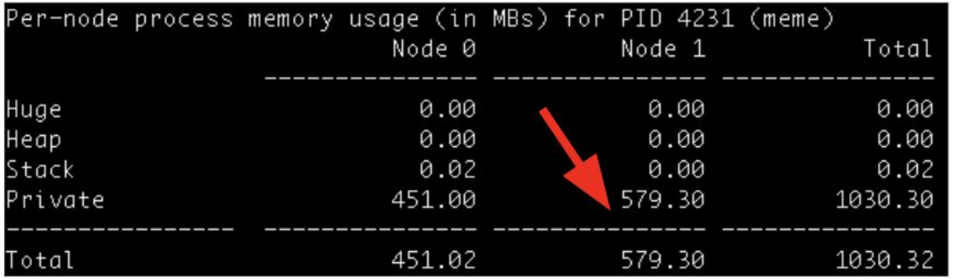


```
Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                   1018.82           11.29         1030.11
----------------  --------------- --------------- ---------------
Total                     1018.84           11.29         1030.12

Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    824.88          205.30         1030.17
----------------  --------------- --------------- ---------------
Total                      824.89          205.30         1030.19

Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    678.94          351.30         1030.23
----------------  --------------- --------------- ---------------
Total                      678.95          351.30         1030.25
```

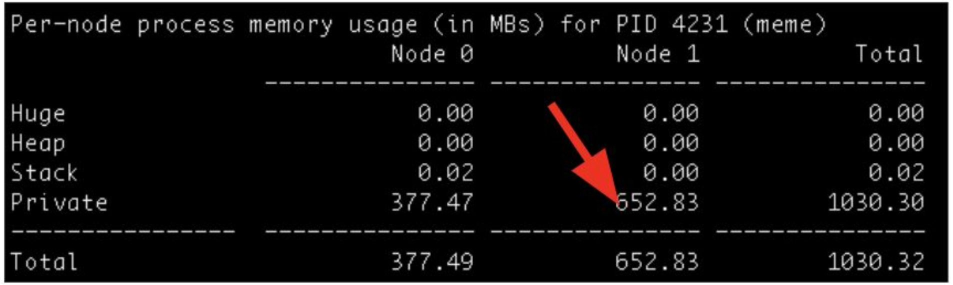
Demo (3)



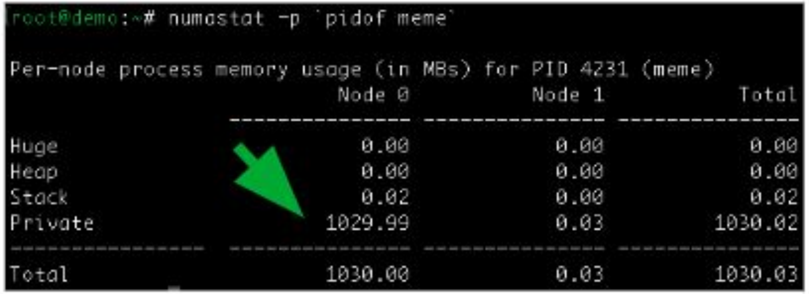
```
Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    575.00          455.30         1030.30
----------------  --------------- --------------- ---------------
Total                      575.02          455.30         1030.31
```



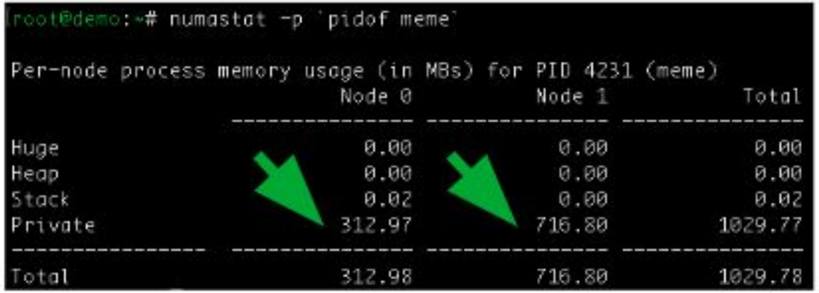
```
Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    451.00          579.30         1030.30
----------------  --------------- --------------- ---------------
Total                      451.02          579.30         1030.32
```



```
Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    377.47          652.83         1030.30
----------------  --------------- --------------- ---------------
Total                      377.49          652.83         1030.32
```



```
root@demo:~# numastat -p `pidof meme`

Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                   1029.99            0.03         1030.02
----------------  --------------- --------------- ---------------
Total                     1030.00            0.03         1030.03
```



```
root@demo:~# numastat -p `pidof meme`

Per-node process memory usage (in MBs) for PID 4231 (meme)
                           Node 0          Node 1           Total
                  --------------- --------------- ---------------
Huge                         0.00            0.00            0.00
Heap                         0.00            0.00            0.00
Stack                        0.02            0.00            0.02
Private                    312.97          716.80         1029.77
----------------  --------------- --------------- ---------------
Total                      312.98          716.80         1029.78
```

Thank you!