



Optimizing Linux Performance

Why is Performance Important

- Regular desktop user
 - Not everyone has the latest hardware
 - Waiting for an application to open
 - Application not responding
 - Memory errors
 - Extra kernel paging
- Efficient application simply runs on more systems, is usable by more people

Contd...

- Do more work with fewer resources
- Run on older hardware
- Save users time
- Higher quality of service for more users
- Need to be a detective to find performance issues, not a hardcore programmer

Linux

- Unprecedented access to source code v/s proprietary windows
- Has powerful performance tools
 - Not centralized however
 - Some are distro standardized
 - Most tools are freelanced
 - Specific tool for specific task
 - Multiple tools for same task

Performance Investigation

- Find a metric, baseline, and target
 - What is the current system performance?
 - What should it be?
- Webserver example
 - How many page requests per second does it serve?
 - How many should it be serving?
- Find a similar configured system that performs well and compare

Contd...

- Track down the approximate problem
- See whether problem has already been solved
 - Bug report results
 - Mailing lists
 - Email the developer

Contd...

- Isolate the problem
- Run same performance tool on a good system and yours and compare
- Change one thing at a time
- Have a changelist with results
- Always re-measure after optimizing
 - Make sure no new problems crop up

Performance information: CPU

- Run Queue statistics
 - Processes competing for CPU time
 - I/O blocks
 - Load average
- Context switches
 - Kernel scheduling
- Interrupts
 - /proc/interrupts
- CPU Utilization

Performance Tools: CPU

- Vmstat (Virtual Memory Statistics)
 - # processes, CPU utilization, # interrupts, # context switches

```
root@ipmiha root]# vmstat
procs          memory          swap          io          system          cpu
r  b  w  swpd  free  buff  cache  si  so  bi  bo  in  cs  us  sy  id
0  0  0  31532 23364 165988 139408  0  0  1  24  8  4  20  2  24
root@ipmiha root]#
```

Oprofile:Details of CPU utilization

- Top:Swiss army knife of system monitoring

```
root@ipmiha root]# top
11:52:10 up 4 days, 14:10, 4 users, load average: 0.02, 0.07, 0.02
113 processes: 111 sleeping, 1 running, 0 zombie, 0 stopped
CPU states:  0.4% user,  0.4% system,  92.2% nice,  0.0% iowait, 16.1% idle
CPU states:  0.4% user,  0.4% system,  0.0% nice,  0.0% iowait, 96.2% idle
Mem:  513204k av, 491096k used,  22100k free,  6484k in_c, 166392k buff
Swap: 1044216k av,  31796k used, 1012420k free

PID USER      PRI  NI  VIRT  RES  SHR  S  T  CPU  MEM  TIME  CPU  COMMAND
22827 yudhan    20  0  1228  828  1100  S  R  0.2  0.2  219  1  xulerc2
22476 root      15  0  44204 34K  1000  S  R  0.3  4.9 1776  1 2
22816 yudhan    15  0  1944 1940  912  S  R  0.3  0.3  6131  1 maglodev
23125 root      15  0  1516 1516  816  S  R  0.3  0.3  0100  0 top
1 root      15  0  464  440  416  S  R  0.0  0.0  0100  1 init
2 root      RT  0  0  0  0  0  S  R  0.0  0.0  0100  0 migration/0
3 root      RT  0  0  0  0  0  S  R  0.0  0.0  0100  1 migration/1
4 root      15  0  0  0  0  S  R  0.0  0.0  0100  0 kworker/
5 root      34 19  0  0  0  S  R  0.0  0.0  0100  0 kworker/0
6 root      34 19  0  0  0  S  R  0.0  0.0  0100  1 kworker/0
11 root     20  0  0  0  0  S  R  0.0  0.0  0100  1 kworker/
7 root      15  0  0  0  0  S  R  0.0  0.0  0116  1 kswapd
8 root      15  0  0  0  0  S  R  0.0  0.0  0100  0 kworker/
```

- Other top options:
 - d = Delay
 - n = iterations
 - i = do not display processes that are not using any of the cpu
- Procinformo – cleaner output than vmstat

```
root@ipmiha root]# procinfo
Linux 2.4.20-8smp (bhcompile@perky) (gcc 3.2.2 20030222) #1 2CPU [ipmiha]
Memory:      Total      Used      Free      Shared  Buffers  Cached
Mem:         513204    498832    24372    0        166400   139360
Swap:        1044216    31868    1012348

Bootup: Thu Feb  2 21:24:36 2006      Load average: 0.87 0.52 0.20 2/135 24612

user  : 1d 15:37:16.38 17.9% page in : 1172662 disk 1: 190014r 941139w
nice  : 5:09:03.84 2.3% page out: 18874748
system: 4:46:01.80 2.1% swap in : 428
idle  : 7d 3:44:52.84 77.6% swap out: 17636
uptime: 4d 14:38:37.42 context :390837239

irq 0: 39831743 timer          irq 10: 0 usb-ohci
irq 1: 18553 keyboard        irq 12: 461603 PS/2 Mouse
irq 2: 0 cascade [4]         irq 14: 5347614 ide0
irq 3: 7                      irq 20: 667848 eth1
irq 4: 6                      irq 21: 1825010 eth0
irq 5: 4                      irq 26: 1126409 aic7xxx
irq 8: 1 rtc                  irq 27: 15 aic7xxx
```

Other procinfo options: Procinfo

- f runs in full screen
- d displays statistics change between samples rather than totals
- n sec Number of seconds to stop between each sample

- Gnome-system-monitor (graphical)
- Mpstat (Multiprocessor statistics)

```
[root@ipmiha root]# mpstat
Linux 2.4.20-8smp (ipmiha) 02/07/2006
12:08:42 PM CPU %user %nice %system %idle intr/s
12:08:42 PM all 17.93 2.33 2.15 77.59 123.72
[root@ipmiha root]#
```

options

- Mpstat -p {cpu | all}
- Mpstat delay

- Sar (System Activity Reporter)

Stores information in a binary file to be replayed later

```
[root@ipmiha root]# sar
Linux 2.4.20-8smp (ipmiha) 02/07/2006
12:00:00 AM CPU %user %nice %system %idle
12:10:00 AM all 18.48 0.30 1.68 79.53
12:20:00 AM all 47.47 0.04 1.48 51.01
12:30:00 AM all 30.28 0.03 0.61 68.88
12:40:00 AM all 14.94 0.40 1.72 82.94
12:50:00 AM all 34.63 0.01 0.93 64.42
01:00:00 AM all 1.90 0.17 1.35 96.50
01:10:00 AM all 3.33 0.10 1.50 95.07
01:20:00 AM all 1.61 0.48 0.83 97.08
01:30:00 AM all 15.29 0.14 1.84 82.72
01:40:00 AM all 30.68 0.06 1.57 67.69
01:50:01 AM all 7.10 0.35 1.97 90.58
02:00:01 AM all 14.03 5.13 2.40 78.44
02:10:00 AM all 5.16 0.78 1.13 92.93
02:20:00 AM all 28.52 5.01 1.97 63.50
02:30:00 AM all 27.44 0.44 0.97 71.14
02:40:00 AM all 26.76 0.51 1.11 71.63
02:50:00 AM all 3.34 1.81 1.26 93.58
03:00:00 AM all 12.51 0.02 0.82 86.65
03:10:00 AM all 38.72 0.37 1.61 59.29
03:20:00 AM all 25.34 0.48 1.46 72.71
03:30:00 AM all 1.92 0.02 0.82 97.24
03:40:00 AM all 6.90 1.54 1.65 89.91
03:50:00 AM all 20.02 3.93 1.73 74.32
04:00:00 AM all 5.41 6.50 1.58 86.51
```

Performance Information: Memory

- Memory subsystem (Virtual memory)
- Swap
 - Total amount of memory currently swapped to disk.
- Buffers and Cache
 - Too Much physical memory
- Active v/s Inactive memory
 - Managing paging
- Processor and high v/s low memory
 - 32 bit addressing more than 1 gb of memory

Performance Tools: Memory

- Vmstat
 - Swap information, memory free and used
- Top
 - Run time toggles sort via memory
- Procfinfo
 - Paging information
- Free
 - How is the system using memory

```
[root@ipmiha root]# free
              total        used         free       shared    buffers     cached
Mem:          513204      484340         28864           0        166116      137284
-/+ buffers/cache:    180940      332264
Swap:        1044216         31828      1012388
[root@ipmiha root]#
```

- Slabtop
 - Kernel allocated caches and how full they are
- Sar
- /proc/meminfo

```

total:      used:      free:      shared:    buffers:    cached:
Mem:  525520896 495296512 30224384      0 170414080 149458944
Swap: 1069277184 32591872 1036685312
MemTotal:      513204 kB
MemFree:       29516 kB
MemShared:      0 kB
Buffers:       166420 kB
Cached:        137000 kB
SwapCached:    8956 kB
Active:        364540 kB
ActiveAnon:    134244 kB
ActiveCache:   230296 kB
Inact_dirty:    36 kB
Inact_laundry: 66240 kB
Inact_clean:   7488 kB
Inact_target:  87660 kB
HighTotal:     0 kB
HighFree:      0 kB
LowTotal:      513204 kB
LowFree:       29516 kB
SwapTotal:    1044216 kB
SwapFree:     1012388 kB

```

Performance Information: Process specific CPU

- Application runtime
 - Kernel or user or library routines
 - Application profiling
- Kernel time v/s user time
- Library time v/s application time
- Subdividing application time

- Ps
 - Dynamic process information

```
[root@ipm1ha root]# ps
  PID TTY          TIME CMD
 20692 pts/3    00:00:00 bash
 29353 pts/3    00:00:00 ps
[root@ipm1ha root]#
```

- Other options
 - -A all options
 - -e all processes
 - r only running processes etc...

Performance Tools: Process Specific CPU

- Time
 - Can be used with the application switch
- Strace
 - Trace system calls of an application
- Ltrace
 - Trace library calls of an application
- Gprof
 - Profile applications (overhead issue, needs to recompile)
- Oprofile
 - Profile applications – process level sampling

Performance Information: I/O

- Total amount and type (read/write) of disk I/O on a system
- Which devices are servicing most of the disk
- Effectiveness of a disk responding to I/O requests
- Which processes are using a given set of files

Performance Tools: I/O

- Vmstat
 - Block/sector read/write information
- Iostat
 - Per device, per partition breakdown
- Sar
 - Lesser functionality than iostat but can simultaneously record many different types of statistics
- Lsof
 - Which processes use which file

Performance Information: Network

- Speed and duplex settings of interfaces
- Amount of network traffic over each interface
- Types of IP traffic flowing in and out of the system
- Amount of each type of IP traffic
- Which applications are generating IP traffic

Performance Tools: Network

- Mii-tool and ethtool - Interface configuration information
- Ifconfig - Configuration and Packet information
- Ip - Performance statistics
- Sar - Detailed network information
- Gkrellm - Graphical real time monitor
- Ipraf - Specific monitoring
- Netstat - Sockets and packet information
- Etherape - Network connection information

Optimizing an application

- Memory
- Is memory usage a problem?
 - Top or ps
 - What type of memory is it using?
 - Cat /proc/<pid>/status
 - Large VmStk? Use gdb and find out which function is using most of the stack
 - Large VmExe? Use nm and remove large symbol code
 - Large VmLib? Use /proc/<pid>/map and streamline large library usage
 - Can also use memory profiler of the programming language

Contd...

- Start up time
- Is application start up time slow?
 - Is loader the problem? Use ld and reduce the number and size of libraries linked
 - Large CPU usage? Use gprof or oprofile to measure critical functions repeatedly
 - Add switches to speed up after looking at gprof information or edit code

Contd...

- I/O
- Is the disk usage a problem?
 - Run iostat and look for overload
 - Spread these files to multiple disks
 - Which function is causing large amount of disk I/O?
 - Which files are accessed most? Use strace and track system calls, spread this evenly between disks or a faster disk

Contd...

- Network
- Is network usage a problem?
 - Use ethtool and check the interface limit
 - Use iptraf and look at the traffic
 - Use ifconfig to check for packet errors
 - Use netstat and find out which application is using that port and generating traffic
 - Use strace and find out which application socket is responsible for the traffic