Hidden Performance Tooling

Marc Skinner

What do mean by Peformance?

- **Is my system slow?**
- **What is my system doing?**
- **Do I have a bottleneck?**
- How has my system been performing – trends?
- What could be the peak performance – benchmarking!
- Monitoring?
- Threshold alerting

Performance tooling in RHEL

- Tools we all use today
  - top
  - vmstat
  - sar
  - iostat
  - free
  - snmp
  - ethtool
  - tuned-adm

Hidden tooling in RHEL7

- Networking
- CPU
- Disk
- Multi-functional

iptraf-ng [NETWORKING]

- TUI to analyze
  - Network performance
  - Statics
  - Flows
- Quick tool to use

nc [NETWORKING]

- nc = NetCat (RHEL6 = nc, RHEL7 = nmap-ncat)
- Useful for testing bandwidth performance of 2 devices – point A to point B

- **Node 1 (point A)**

```
# nc -l -n 12345 > /dev/null
```

- **Node 2 (point B :: 1gb NIC)**

```
# dd if=/dev/zero bs=1M count=1024 | nc -n 192.168.33.13 12345
1024+0 records in
1024+0 records out
1073741824 bytes (1.1 GB) copied, 53.3481 s, 20.1 MB/s
```

- **Node 3 (point C :: 1gb NIC)**

```
#dd if=/dev/zero bs=1M count=1024 | nc -n 192.168.33.44 12345
1024+0 records in
1024+0 records out
1073741824 bytes (1.1 GB) copied, 9.11378 s, 118 MB/s
```

qperf [NETWORKING]

- Tool to measure bandwidth and latency between two nodes
- **Node 1 (point A :: 1gb NIC)**

qperf

- **Node 2 (point B :: 1gb NIC)**

qperf 192.168.33.13 tcp_bw tcp_lat

tcp_bw:

bw = **19.2** MB/sec

tcp_lat:

latency = 276 us

qperf 192.168.33.13 udp_bw udp_lat

udp_bw:

send_bw = **120** MB/sec

recv_bw = 78.2 MB/sec

udp_lat:

latency = 552 us

ibstatus [NETWORKING]

- Infiniband interface status – similar to ethtool for Ethernet interfaces

```
# ibstatus
Infiniband device 'mthca0' port 1 status:
default gid:     fe80:0000:0000:0000:0005:ad00:000c:6de1
base lid:  0x4
sm lid:          0x1
state:           4: ACTIVE
phys state:      5: LinkUp
rate:            20 Gb/sec (4X DDR)
link_layer:      InfiniBand
```

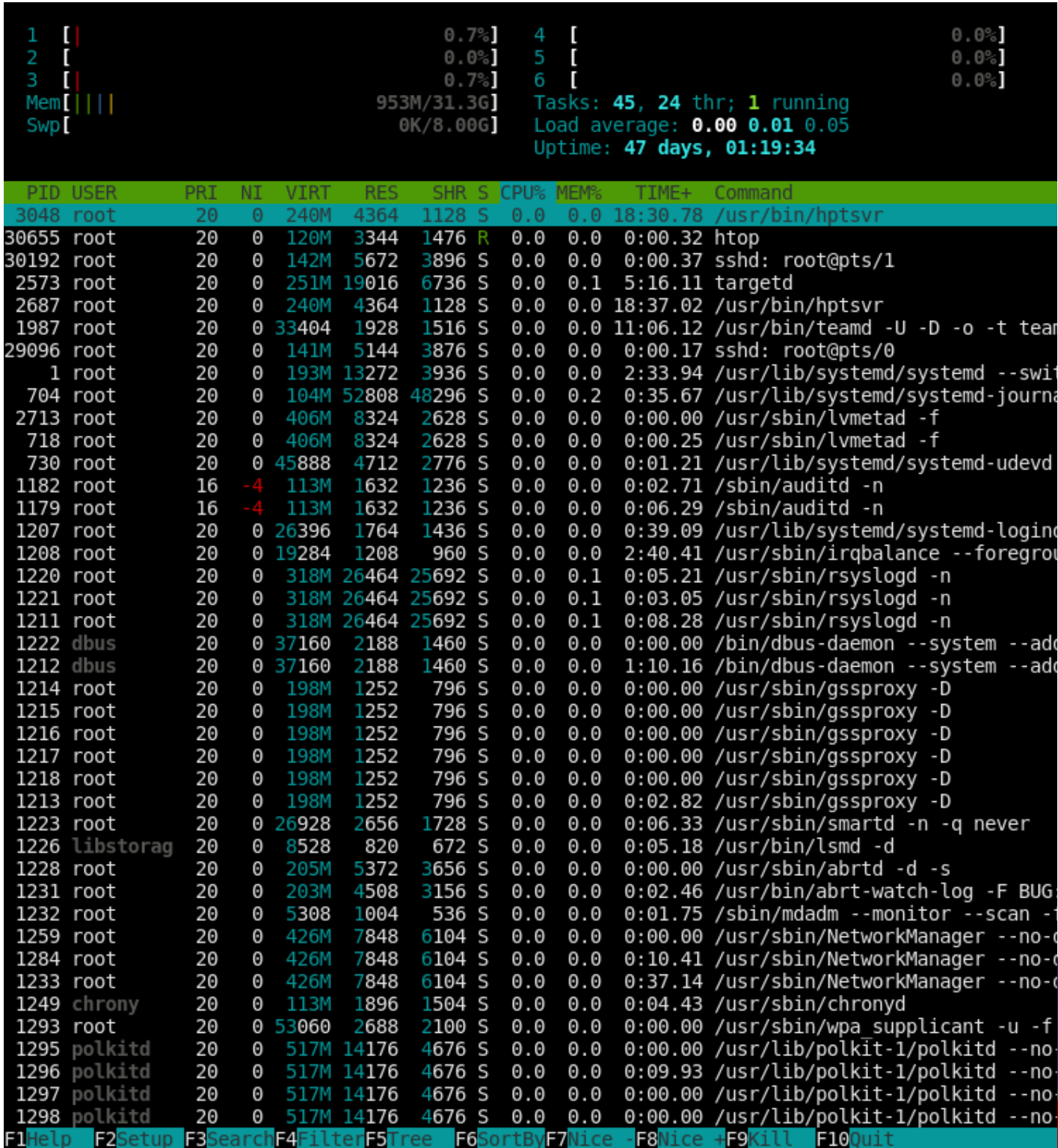
ifstat [NETWORKING]

- Dumps networking statistics

```
[root@iscsi ~]# ifstat
#kernel
Interface        RX Pkts/Rate    TX Pkts/Rate    RX Data/Rate    TX Data/Rate
                 RX Errs/Drop    TX Errs/Drop    RX Over/Rate    TX Coll/Rate
lo                     0 0             0 0             0 0             0 0
                       0 0             0 0             0 0             0 0
enp8s0                 0 0             0 0             0 0             0 0
                       0 0             0 0             0 0             0 0
enp2s0f0             158 0            57 0         19264 0         10879 0
                       0 0             0 0             0 0             0 0
enp2s0f1               0 0             0 0             0 0             0 0
                       0 0             0 0             0 0             0 0
enp9s0               326 0             6 0         32012 0           640 0
                       0 0             0 0             0 0             0 0
enp10s0              198 0             6 0         25312 0           670 0
                       0 0             0 0             0 0             0 0
ib0                10572 0         13372 0          4909K 0       131868K 0
                       0 14            0 0             0 0             0 0
ib1                   14 0             0 0          1232 0             0 0
                       0 14            0 0             0 0             0 0
team0                682 0            55 0         64312 0         10239 0
                       0 0             0 0             0 0             0 0
[root@iscsi ~]#
```

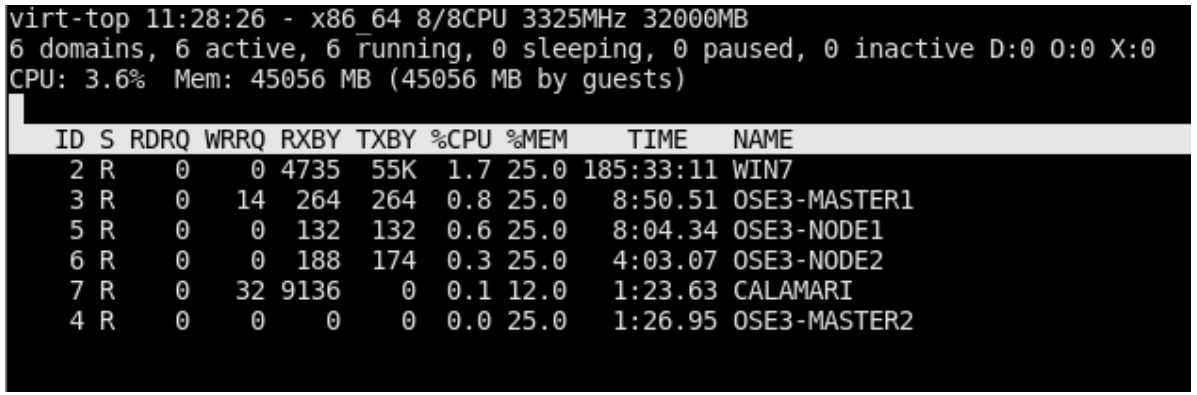
htop [CPU]

- TUI for running processes – lets you scroll horizontally and vertically
- Similar to top



virt-top [CPU]

- Top like tool for virtualized resources when using KVM



dstat [DISK]

- Great tool for viewing vmstat, iostat, ifstat interactively.
- Can build your table by including/excluding many options
- Nice for correlation of multiple resources

```
[root@iscsi ~]# dstat
You did not select any stats, using -cdngy by default.
----total-cpu-usage---- -dsk/total- -net/total- ---paging-- ---system--
usr sys idl wai hiq siq| read  writ| recv  send|  in   out | int   csw
 13   0  86   0   0   0|  25M  908k|   0     0 |   0     0 |1782   906
  0   0 100   0   0   0|2481k   29k|  24k 2500k|   0     0 | 657   564
  0   0 100   0   0   0|2505k   74k|  48k 2522k|   0     0 | 720   636
  0   0 100   0   0   0|1000k   46k|  18k 1005k|   0     0 | 294   323
  0   0 100   0   0   0|   0   127k|  64k 3888B|   0     0 | 267   342
  0   0 100   0   0   0|2028k 2176k|  13k 1008k|   0     0 | 298   270
  0   0 100   0   0   0|8192B  366k| 155k   11k|   0     0 | 231   265
  0   0 100   0   0   0|1004k   31k|  13k 1008k|   0     0 | 205   203
  0   0 100   0   0   0|   0    43k|  23k 2208B|   0     0 | 157   224
  0   0 100   0   0   0|1004k   96k|  47k 1010k|   0     0 | 314   322
  0   0 100   0   0   0|   0    46k|  22k 2208B|   0     0 | 158   214
  0   0 100   0   0   0|1000k   29k|  13k 1004k|   0     0 | 196   201
  0   0 100   0   0   0|   0    90k|  23k 2208B|   0     0 | 197   292
  0   0 100   0   0   0|2481k   29k|  24k 2498k|   0     0 | 674   569
  0   0 100   0   0   0|2505k  116k|  69k 2525k|   0     0 | 802   726
  0   0 100   0   0   0|2796k 3785k|  46k 1009k|   0     0 | 407   406
  0   0 100   0   0   0|4096B   97k|  22k 6464B|   0     0 | 191   243
```

iotop [DISK]

- Simple I/O top tool

```
Total DISK READ :      20.51 M/s | Total DISK WRITE :       0.00 B/s
Actual DISK READ:      20.51 M/s | Actual DISK WRITE:      19.56 M/s
  TID  PRIO  USER     DISK READ  DISK WRITE  SWAPIN     IO>    COMMAND
29221 be/4 root     1001.27 K/s    0.00 B/s  0.00 %  0.00 % [iscsi_trx]
25135 be/4 root     1752.22 K/s    0.00 B/s  0.00 %  0.00 % [kworker/2:2]
28061 be/4 root        5.38 M/s    0.00 B/s  0.00 %  0.00 % [kworker/3:1]
16299 be/4 root        9.29 M/s    0.00 B/s  0.00 %  0.00 % [kworker/4:1]
27103 be/4 root        2.20 M/s    0.00 B/s  0.00 %  0.00 % [kworker/0:2]
 8676 be/4 root      977.80 K/s    0.00 B/s  0.00 %  0.00 % [iscsi_trx]
    1 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % systemd --s~erial
    2 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [kthreadd]
    3 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [ksoftirqd/0]
    7 rt/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [migration/0]
    8 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcu_bh]
    9 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/0]
   10 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/1]
   11 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/2]
   12 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/3]
   13 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/4]
   14 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuob/5]
   15 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcu_sched]
   16 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/0]
   17 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/1]
   18 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/2]
   19 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/3]
   20 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/4]
   21 be/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [rcuos/5]
   22 rt/4 root        0.00 B/s    0.00 B/s  0.00 %  0.00 % [watchdog/0]
```

blktrace [DISK]

- Tool for generating/capturing disk I/O on block devices
- Select your block device, capture duration, output file

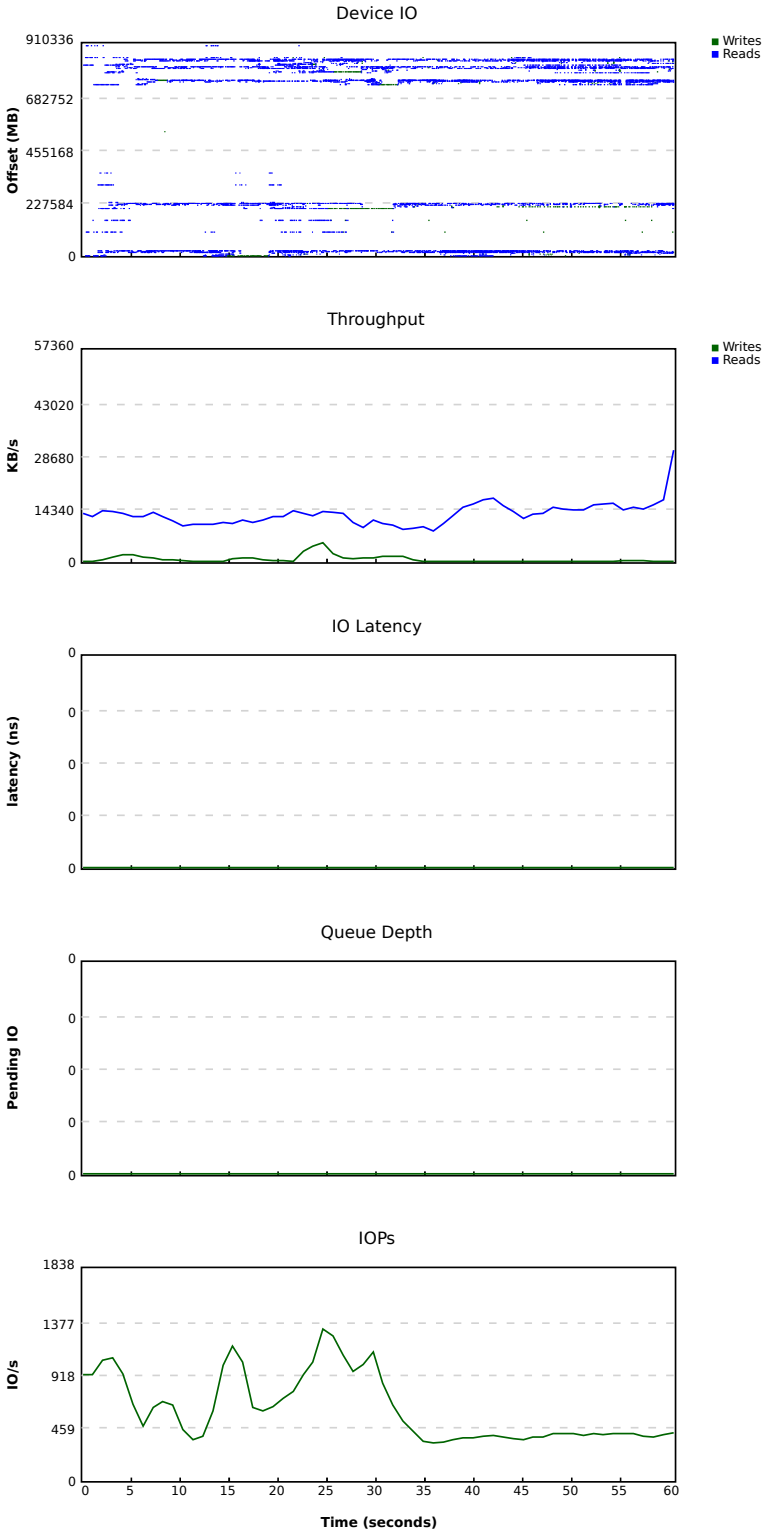
```
# blktrace -d /dev/md4 -w60 -o kvm_images
=== md4 ===
  CPU  0:                  286 events,       14 KiB data
  CPU  1:                  462 events,       22 KiB data
  CPU  2:                  471 events,       23 KiB data
  CPU  3:                  266 events,       13 KiB data
  CPU  4:                   44 events,       3 KiB data
  CPU  5:                   64 events,       4 KiB data
  CPU  6:                   91 events,       5 KiB data
  CPU  7:                   30 events,       2 KiB data
  Total:                 1714 events (dropped 0),      81 KiB data
```

blktrace + iowatcher [DISK]

- Great, I have data but now what?
- blktrace captures data in binary format
- Need to read – graph?

iowatcher -t kvm_images.blktrace -o kvm_images-trace.svg

- Generates a graph from captured data!
- Can even generate movie in ogg or mp4 format for visual playback of your data capture.



Cockpit [Multi-functional]

- Install, enable and start cockpit service on systems

subscription-manager repos --enable rhel-7-server-extras-rpms

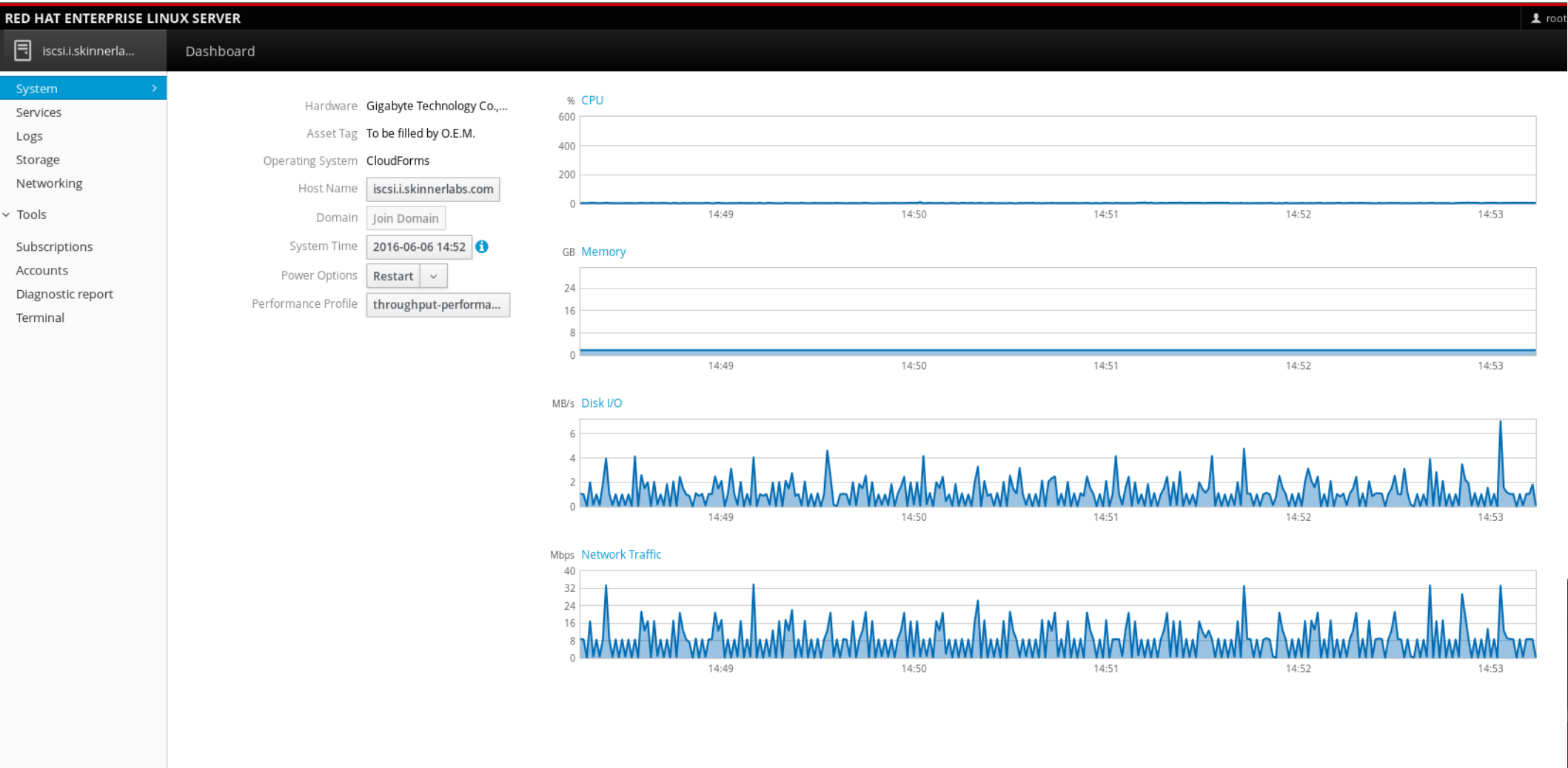
yum install cockpit

systemctl start cockpit

- Connect via port 9090 with web browser
- https://myserver.com:9090
- Enter local username/password

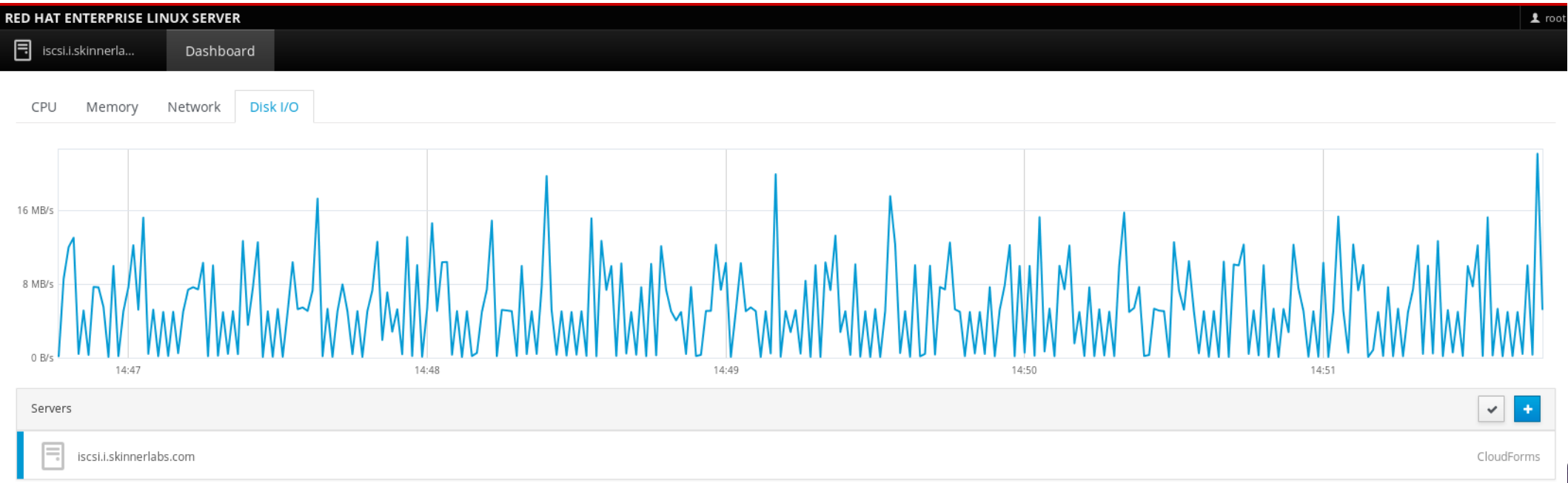
Cockpit :: Server View

- Connect to Cockit agent via web browser on port 9090



Cockpit :: Dashboard

- CPU :: Memory :: Network :: Disk I/O



PCP [Multi-functional]

- PCP = Performance Co-Pilot (RHEL 7 and RHEL => 6.6)
- RHEL 7 how to install:

```
# yum install pcp
# systemctl enable pmcd
# systemctl enable pmlogger
# systemctl start pmcd
# systemctl start pmlogger
```

- Data is collected every 60 seconds by default (sampled)
- To change, edit:
  - /etc/pcp/pmlogger/control
  - Append -t 10s to LOCALHOSTNAME line to change to 10 second interval
  - Restart pmlogger service

PCP :: Charts

- Extensive list of PCP metric chart add-ons

yum search pcp | grep pmda

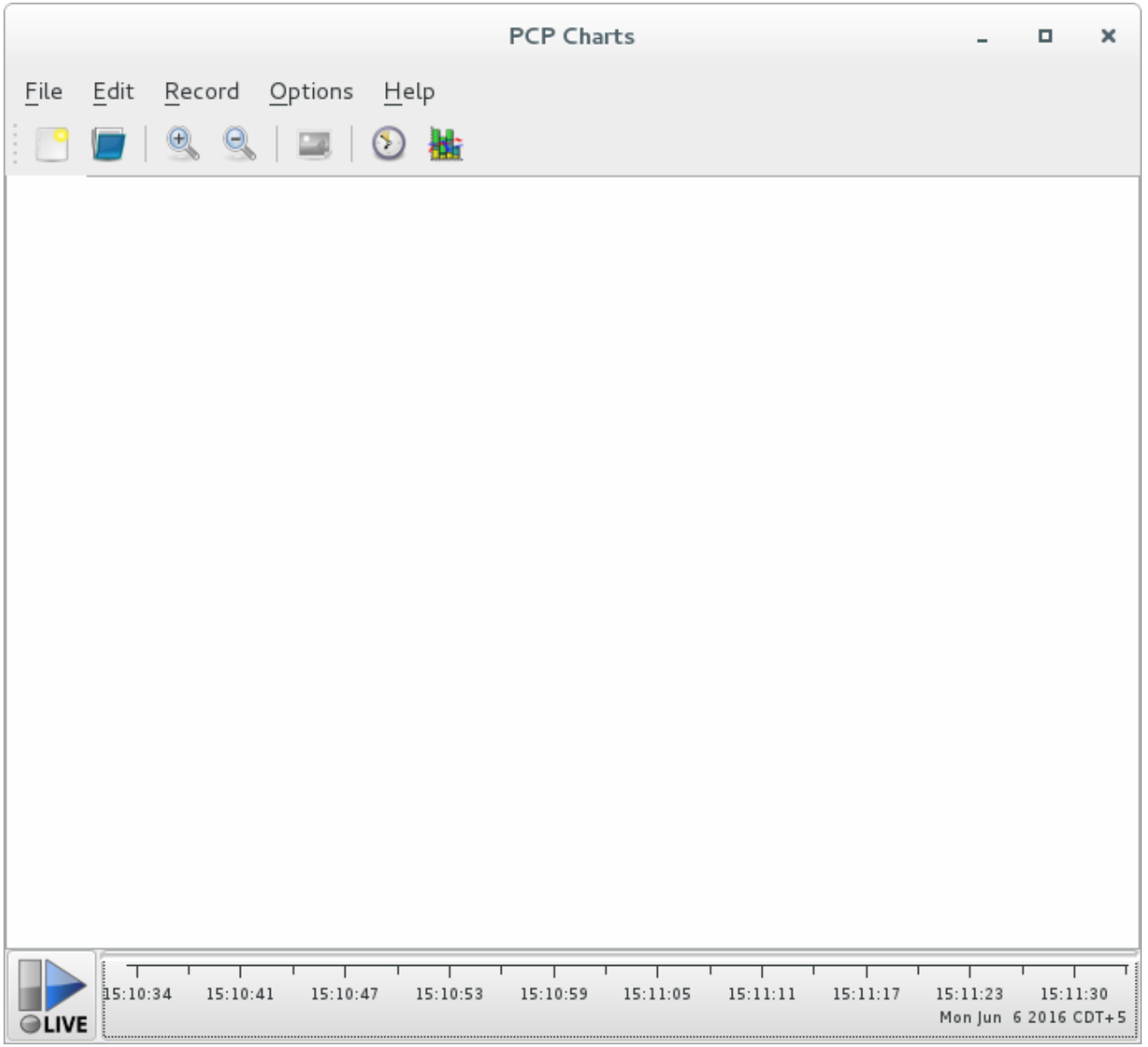
ActiveMQ
Apache
Bash shell
Bonded network
Cifs shell
Cisco shell
Device Mapper
389 Directory
Gfs2 shell
Gluster
GPFS Filesystem
Infiniband
JSON data
KVM
Lmsensors
Arbitrary log
Lustre
Mailq shell
Memcached
Mounts shell
MySQL
Named
Netfilter
Usenet News
NFS Clients
Nginx
Nvidia
Performance API
PowerDNS
Postfix
PostgreSQL
Roomtemp
Rpm shell
Rsyslog
Samba
Sendmail
Simple Network
Systemd
Trace shell
Unbound DNS
VMware
Weblog shell
Zimbra

PCP :: Client/Server

- Client: pmchart
- Server: pmcd and pmlogger

PCP :: Client

- GUI client interface

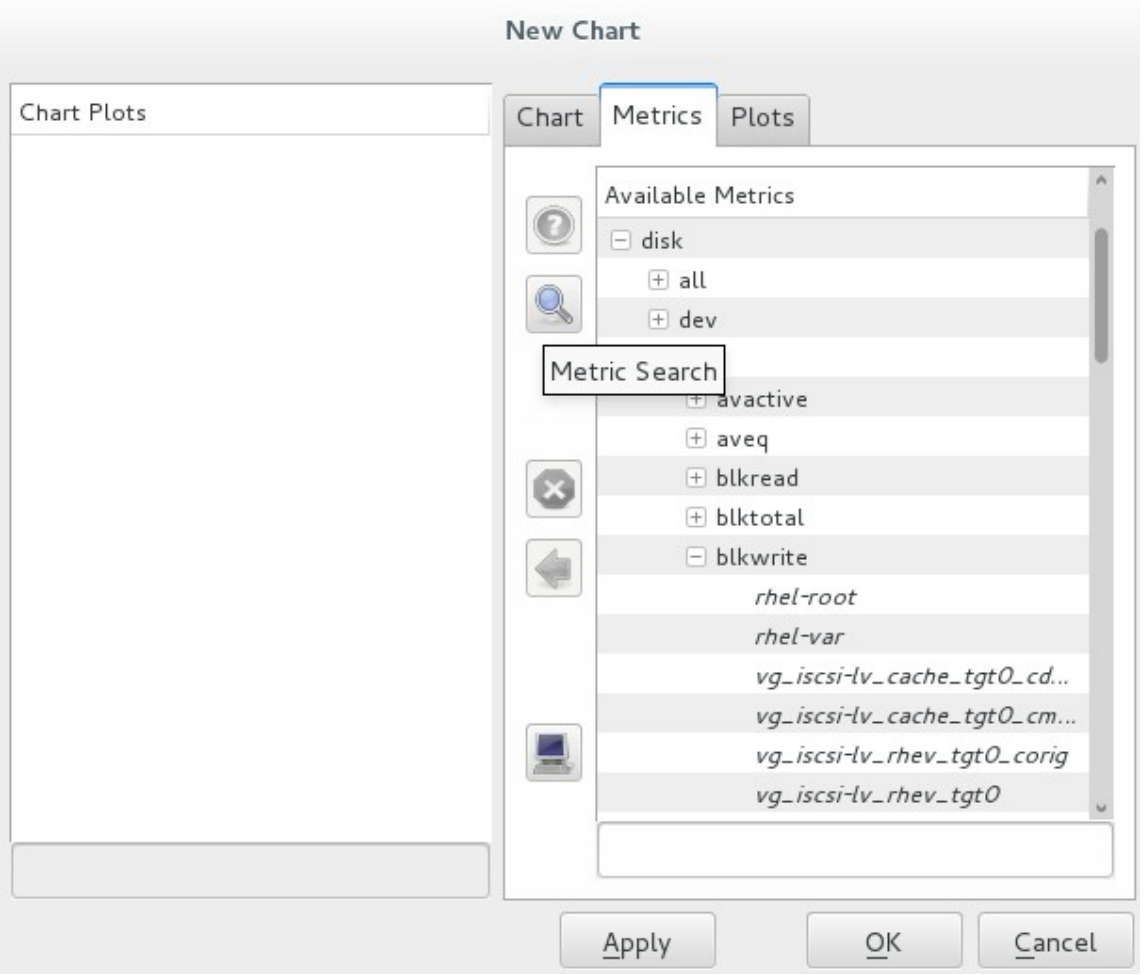


PCP :: Client

- Open pre-configured charts
- Select from Open View list

PCP :: Client

- Create new charts
- Choose host to monitor
- Select from available metrics
- Select 1 to n metric

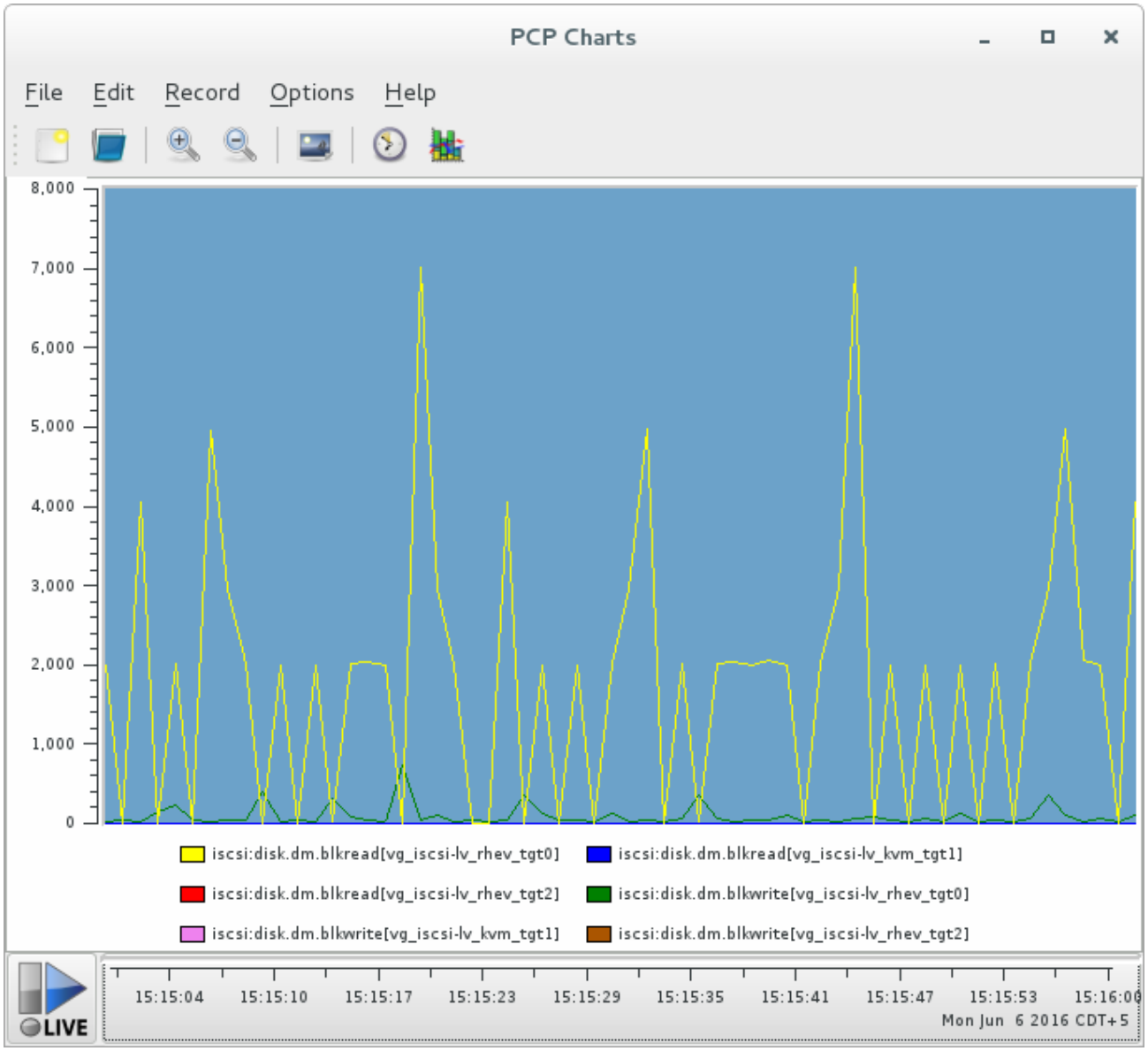


PCP :: Client

- Disk chart created
- Disk blkread
- Disk blkwrite

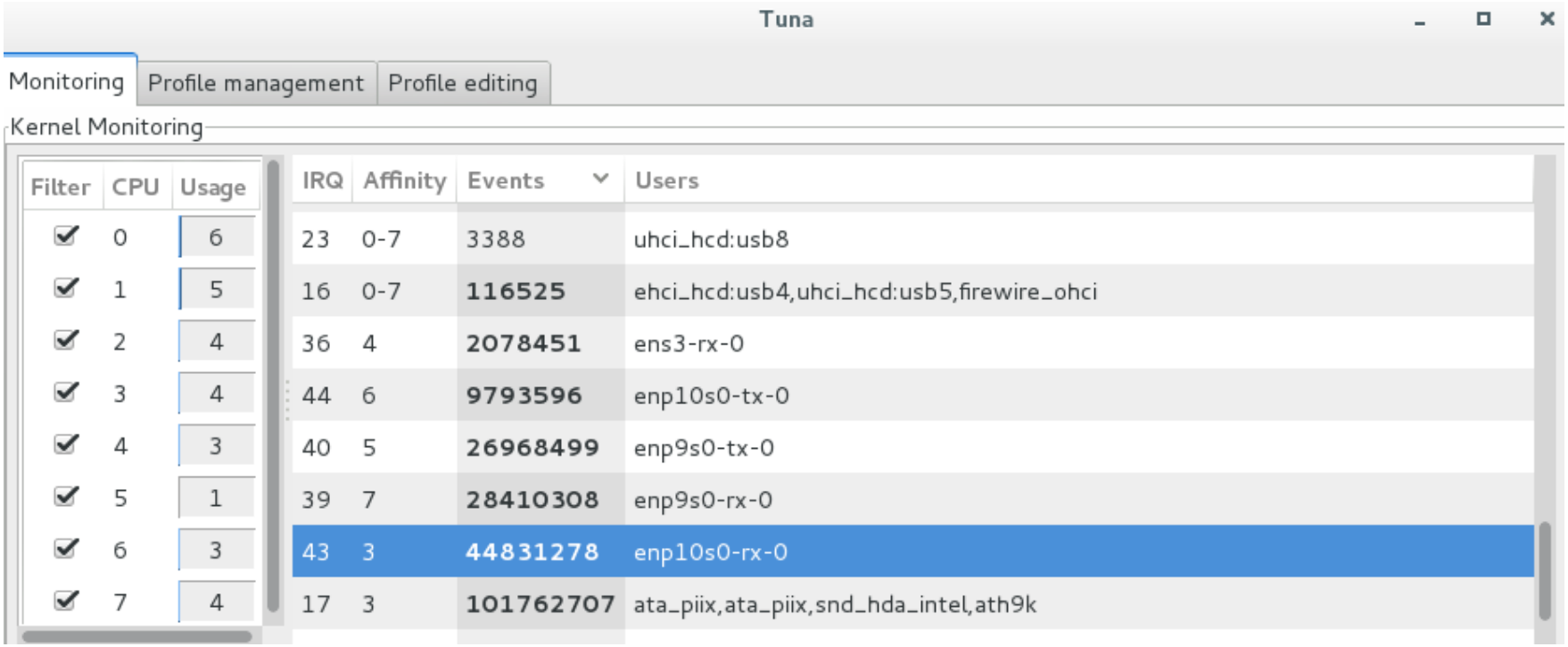
- Export as graphic
- Record for playback

- Import data from Collectd
- Export data to Webapps
  - Grafana
  - Graphite
  - Vector



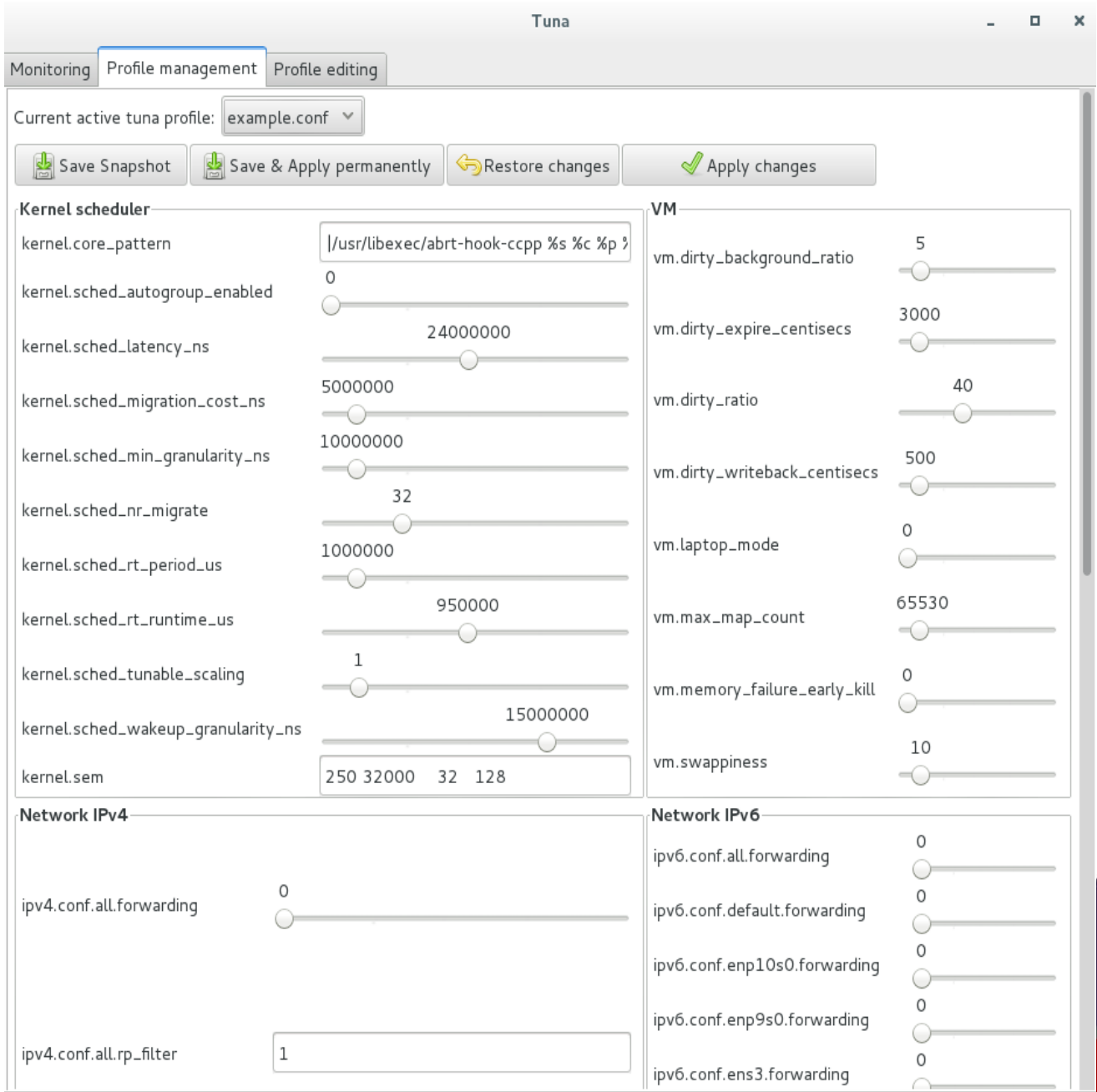
tuna [Multi-functional]

- Monitor CPU/IRQ affinity
- Make changes, tweak
- RX/TX pinned to unique cores/threads



tuna [Multi-functional]

- Create a profile
- Change in real time
- Be careful – loaded gun!
- Save for later



redhat.

Questions?

What are you using?