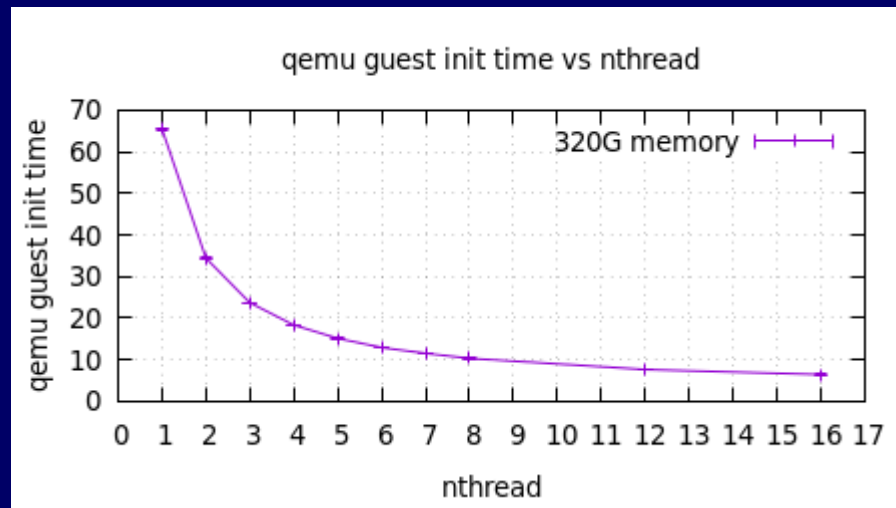# ktask: multithread CPU-intensive kernel work

- Problem: A single thread can spend excessive time in the kernel
  - Often for init- and fini-related jobs that scale with system size
  - 60 seconds to start 320G VFIO-enabled kvm guest
  - 3.9 seconds to initialize struct pages at boot on a 375G node

- Solution: parallelize!

thr	speedup	average sec
1		65.5 (± 0.4%)
2	1.9x	34.4 (± 0.3%)
3	2.8x	23.7 (± 0.2%)
4	3.6x	18.3 (± 0.3%)
5	4.3x	15.1 (± 0.3%)
6	5.1x	12.9 (± 0.1%)
7	5.7x	11.5 (± 0.6%)
8	6.4x	10.3 (± 0.8%)
12	8.6x	7.6 (± 0.6%)
16	10.2x	6.4 (± 0.6%)

qemu guest init time vs nthread

Daniel Jordan, Oracle Core Kernel Team

ktask: multithread CPU-intensive kernel work

- ktask features
  - divides and load balances the job
  - helper threads run at MAX_NICE to avoid system disturbance
  - executes job on local node, given node(s), or any node
  - per-node and system-wide caps on number of helper threads
  - cgroup-aware (coming soon)

```
  PID USER      PRI  NI  VIRT   RES   SHR S CPU% MEM%   TIME+  Command
 4182 root       20   0  224G  179G   132 S 692. 23.8  0:35.17 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4186 root       20   0  224G  179G   132 R  99. 23.8  0:05.05 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4188 root       20   0  224G  179G   132 R  99. 23.8  0:05.04 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4189 root       20   0  224G  179G   132 R 98.6 23.8  0:05.04 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4187 root       20   0  224G  179G   132 R 98.6 23.8  0:05.03 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4199 root       20   0  102G 20.2G  6296 R 98.6  2.7  0:03.90 /usr/bin/qemu-system-x86_64 -name vmol74 -machine pc-q35-2.11,accel=kv
 4184 root       20   0  224G  179G   132 R 98.6 23.8  0:05.00 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4183 root       20   0  224G  179G   132 R 98.6 23.8  0:04.98 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 4185 root       20   0  224G  179G   132 R 97.9 23.8  0:04.97 /home/dbbench/src/vm-scalability/usemem -t 7 -j 4096 34359738368
 3865 root       39  19     0     0     0 R  2.0  0.0  0:03.08 kworker/u202:2+ktask_wq
 3536 root       39  19     0     0     0 R  1.3  0.0  0:03.08 kworker/u202:0+ktask_wq
 3870 root       39  19     0     0     0 R  1.3  0.0  0:03.09 kworker/u202:6+ktask_wq
 3867 root       39  19     0     0     0 R  1.3  0.0  0:02.45 kworker/u202:3+ktask_wq
 3864 root       39  19     0     0     0 R  1.3  0.0  0:02.45 kworker/u202:1+ktask_wq
 3868 root       39  19     0     0     0 R  1.3  0.0  0:03.02 kworker/u202:4+ktask_wq
 3869 root       39  19     0     0     0 R  1.3  0.0  0:03.03 kworker/u202:5+ktask_wq
```

ktask: multithread CPU-intensive kernel work

- Open issues

  - How to determine maximum threads per task?
    - Make the client provide it? Determine through testing on representative systems.

  - How to disable ktask on energy-conscious devices?
    - Check the current scaling driver and governors

Daniel Jordan, Oracle Core Kernel Team   Latest patches: https://bit.ly/2DCO6Zb