AutoNUMA bench

Red Hat, Inc.

Andrea Arcangeli
aarcange at redhat.com

git clone --reference linux -b autonuma git://git.kernel.org/pub/scm/linux/kernel/git/andrea/aa.git
6e7267f0c9973f207a826c6b1fdae4e69c54ea80

26 Mar 2012

Copyright © 2012 Red Hat Inc.

redhat.

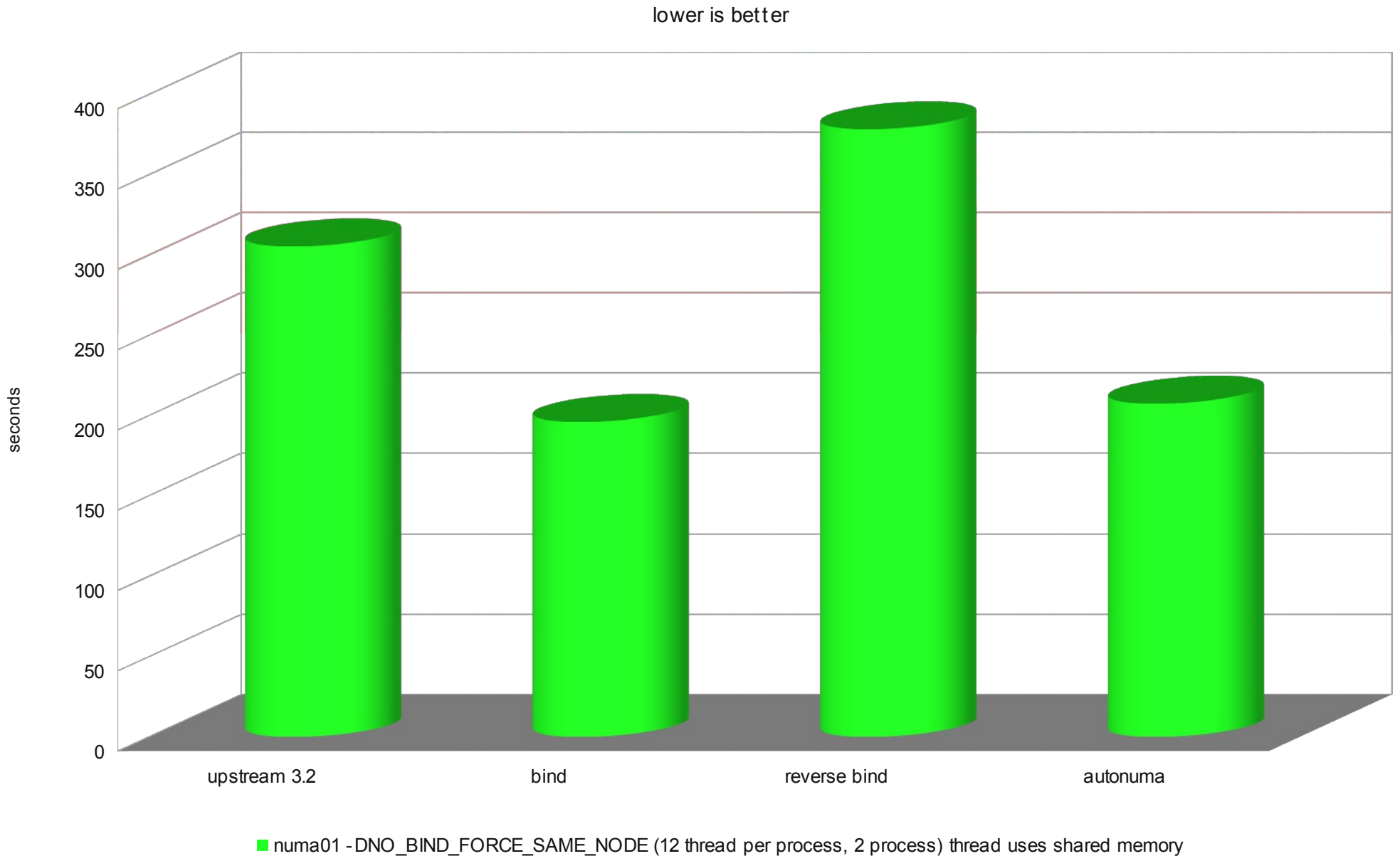
Hardware

- 2 NUMA nodes
- 2 CPU sockets
- 6 CPU cores per socket
- 2 HT CPU threads per core (total 24 CPUs)
- 8GB of RAM per node (total 16 GB of RAM)

Copyright © 2012 Red Hat Inc.

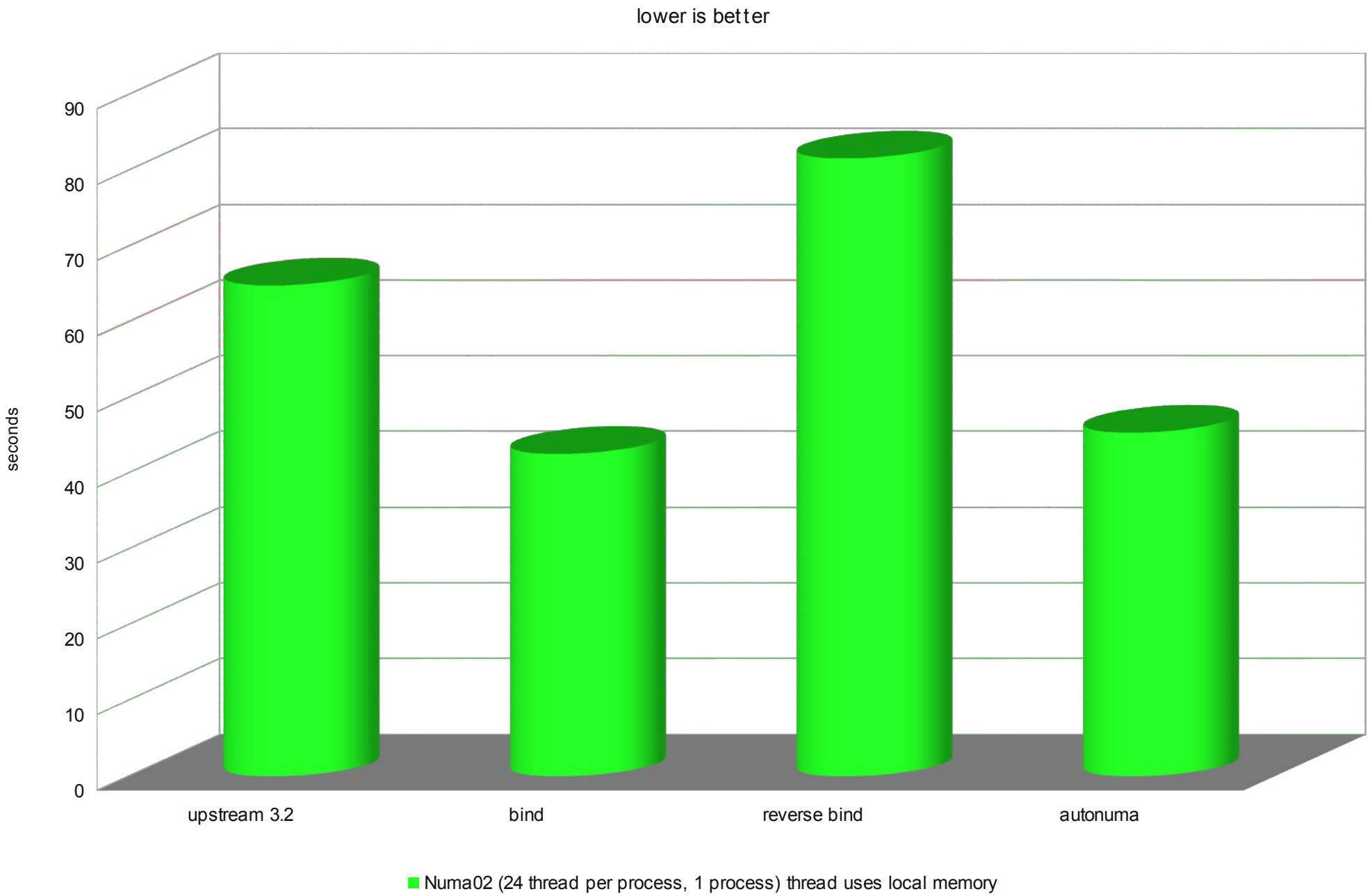
redhat.

numa01 -DNO_BIND_FORCE_SAME_NODE
all threads shares the same memory, 12 threads per process, 2 processes



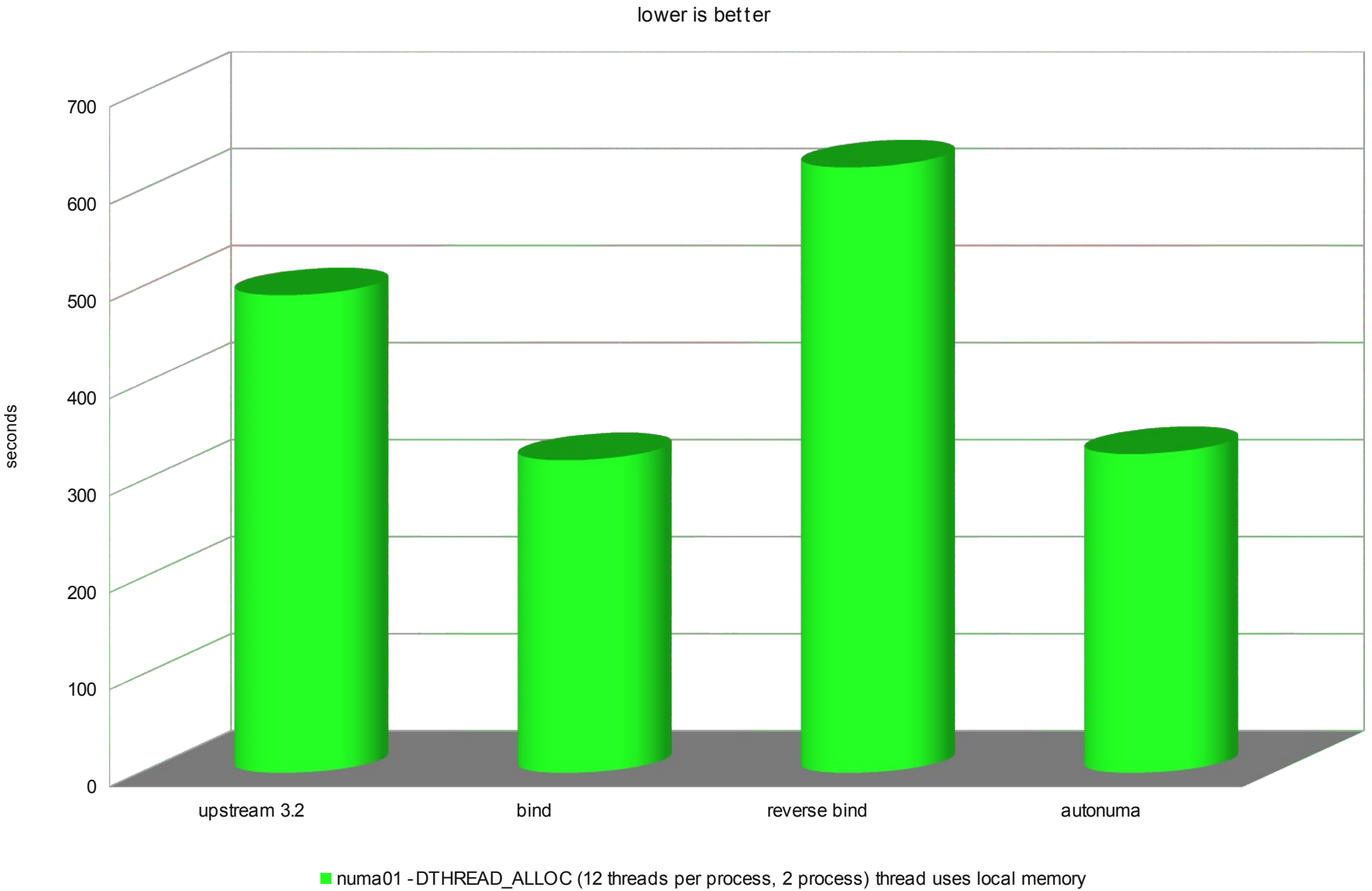
Copyright © 2012 Red Hat Inc.

redhat



Copyright © 2012 Red Hat Inc.

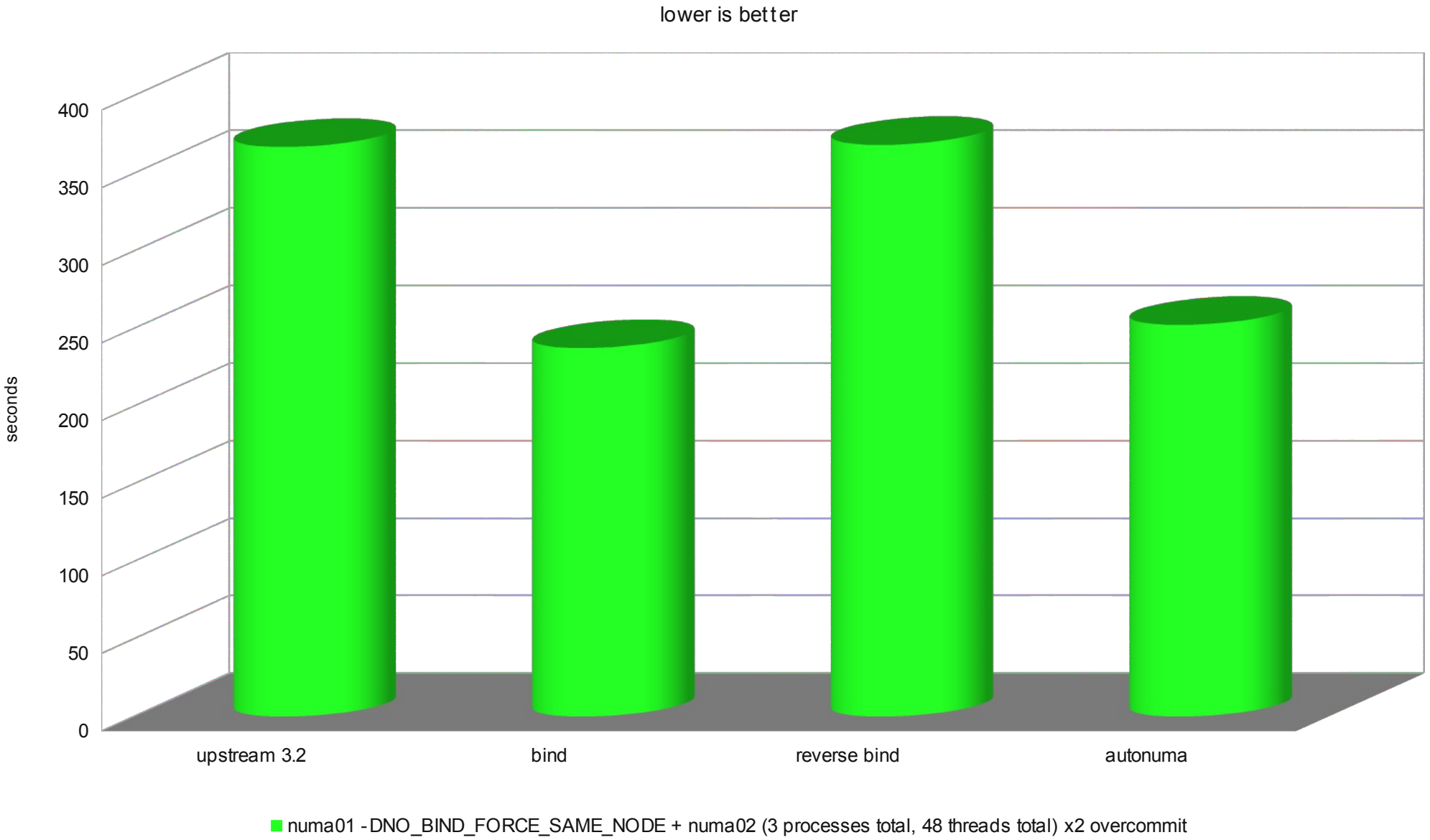
numa01 per-thread local memory, 12 threads per process, 2 processes



Copyright © 2012 Red Hat Inc.

redhat

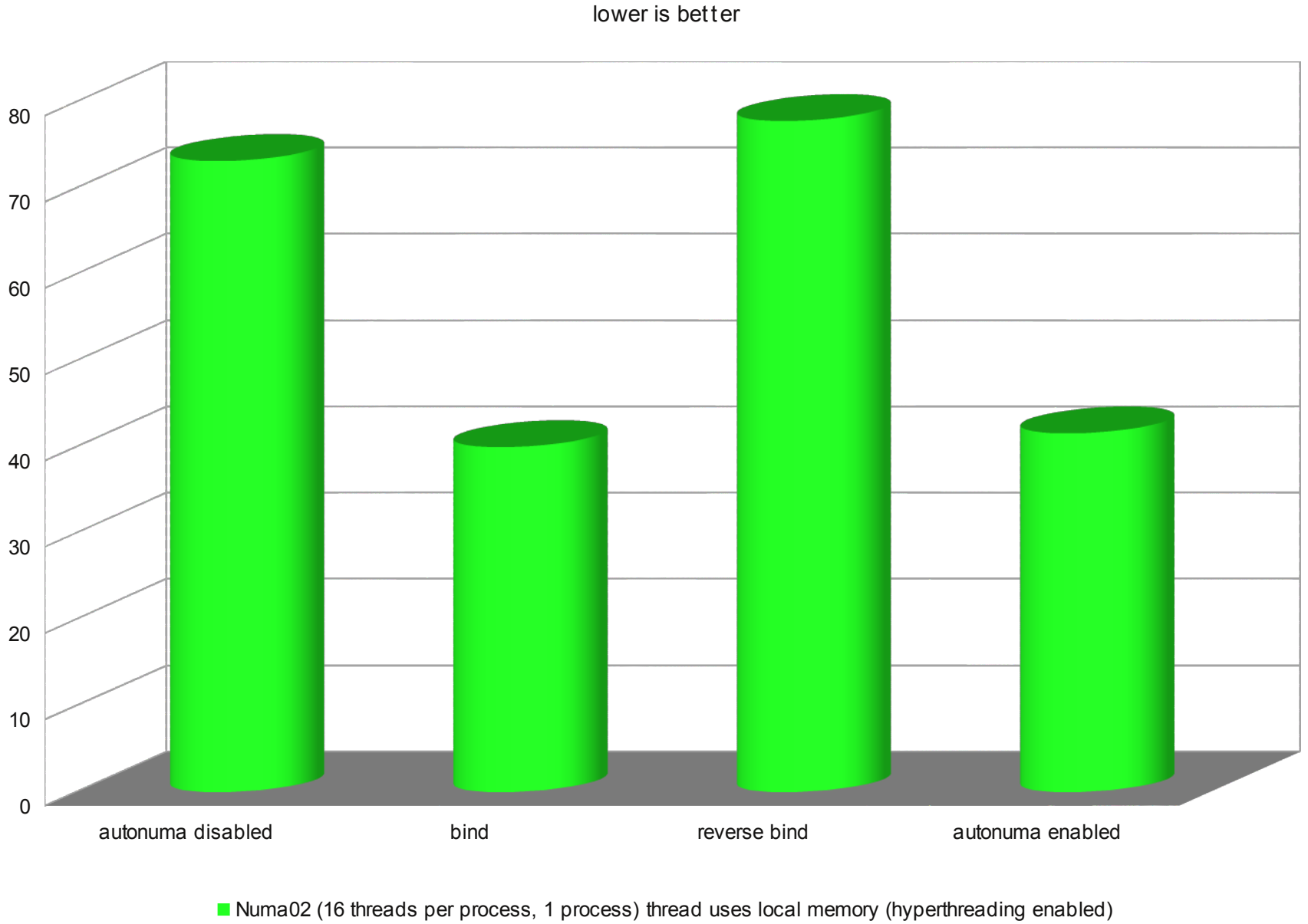
x2 CPU overcommit: numa01 -DNO_BIND_FORCE_SAME_NODE + numa02
24 threads using local memory +
12 threads using shared memory +
12 threads using shared memory



Copyright © 2012 Red Hat Inc.

redhat

numa02 per-thread local memory, 12 threads per process 1 process (HT enabled)
SMT testcase



Copyright © 2012 Red Hat Inc.

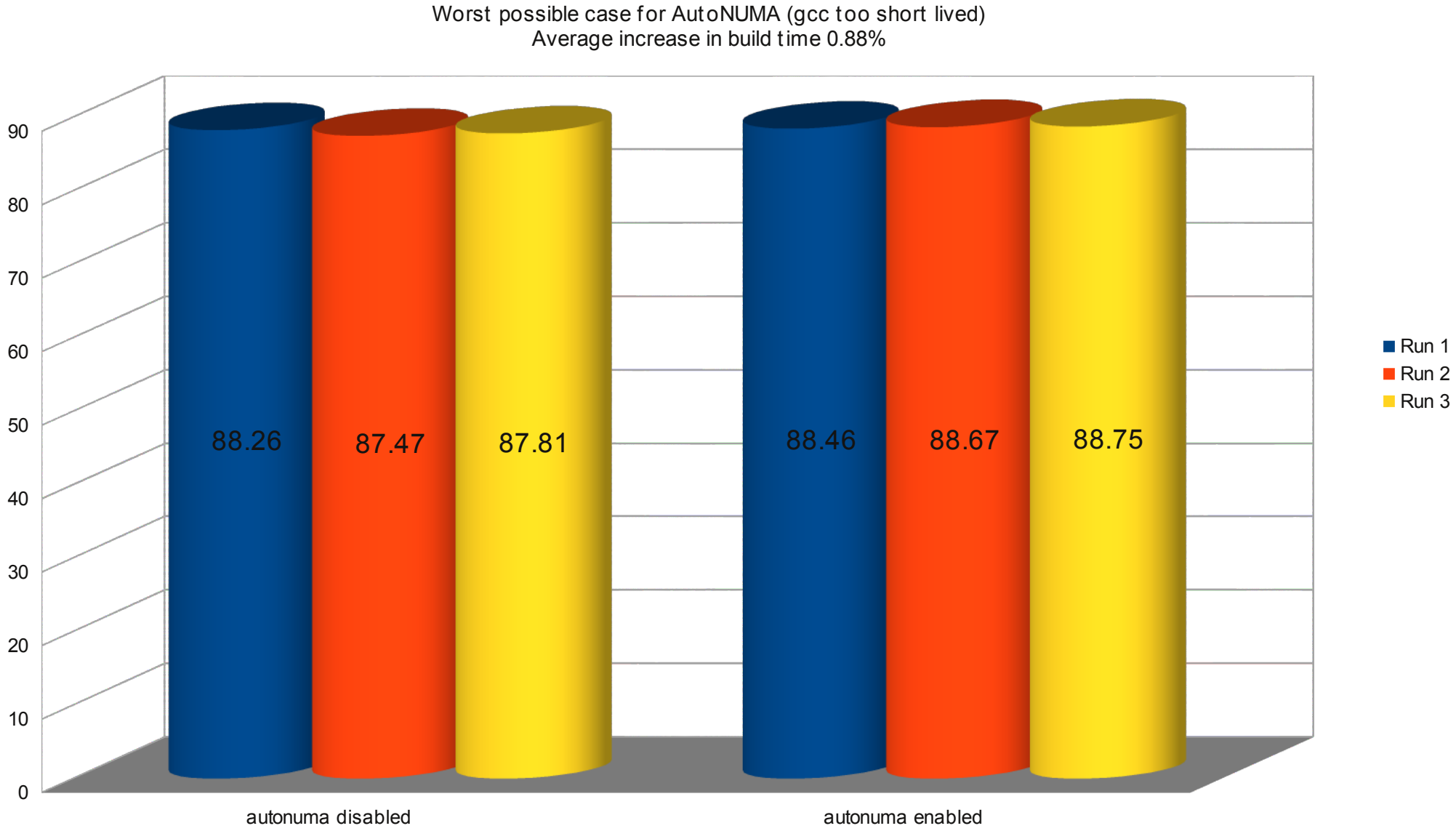
redhat

autonuma benchmark (hash 7e4dc3dbbda23b873ca7771b5cf296078e6ed1f7 vs 3.2 upstream default vs 3.2 upstream bind vs upstream inverse bind)				
	autonuma off	bind	reverse bind	autonuma
numa01 -DNO_BIND_FORCE_SAME_NODE (12 thread per process, 2 process) thread uses shared memory	305.36	196.07	378.34	207.47
Numa02 (24 thread per process, 1 process) thread uses local memory	64.81	42.58	81.6	45.39
numa01 -DTHREAD_ALLOC (12 threads per process, 2 process) thread uses local memory	491.88	321.94	623.62	328.43
numa01 -DNO_BIND_FORCE_SAME_NODE + numa02 (3 processes total, 48 threads total) x2 overcommit	366.96	237.43	368.35	252.31

Autonuma SMT fix uses hash 6e7267f0c9973f207a826c6b1fdae4e69c54ea80 Numa02 (16 threads per process, 1 process) thread uses local memory (hyperthreading enabled)	73.16	39.99	77.8	41.59

Copyright © 2012 Red Hat Inc.

Kernel build time in seconds on tmpfs (make -j32)
Autonuma enabled includes one knuma_scand pass every 10sec



Copyright © 2012 Red Hat Inc.

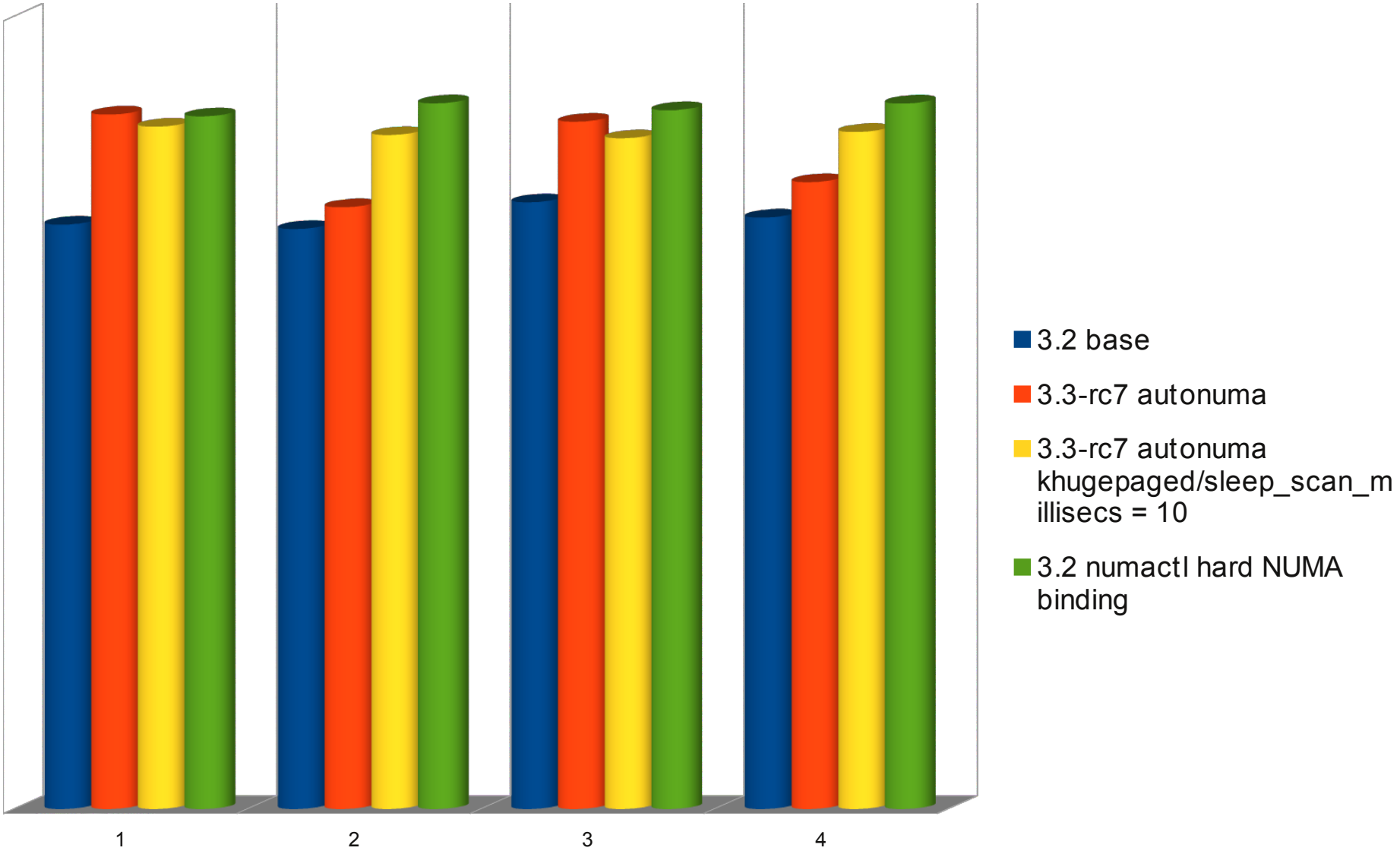
redhat

autonuma overhead kernel build tmpfs (make -j32)	Run 1	Run 2	Run 3
autonuma disabled	88.262	87.465	87.807
autonuma enabled	88.459	88.669	88.745

Copyright © 2012 Red Hat Inc.

redhat.

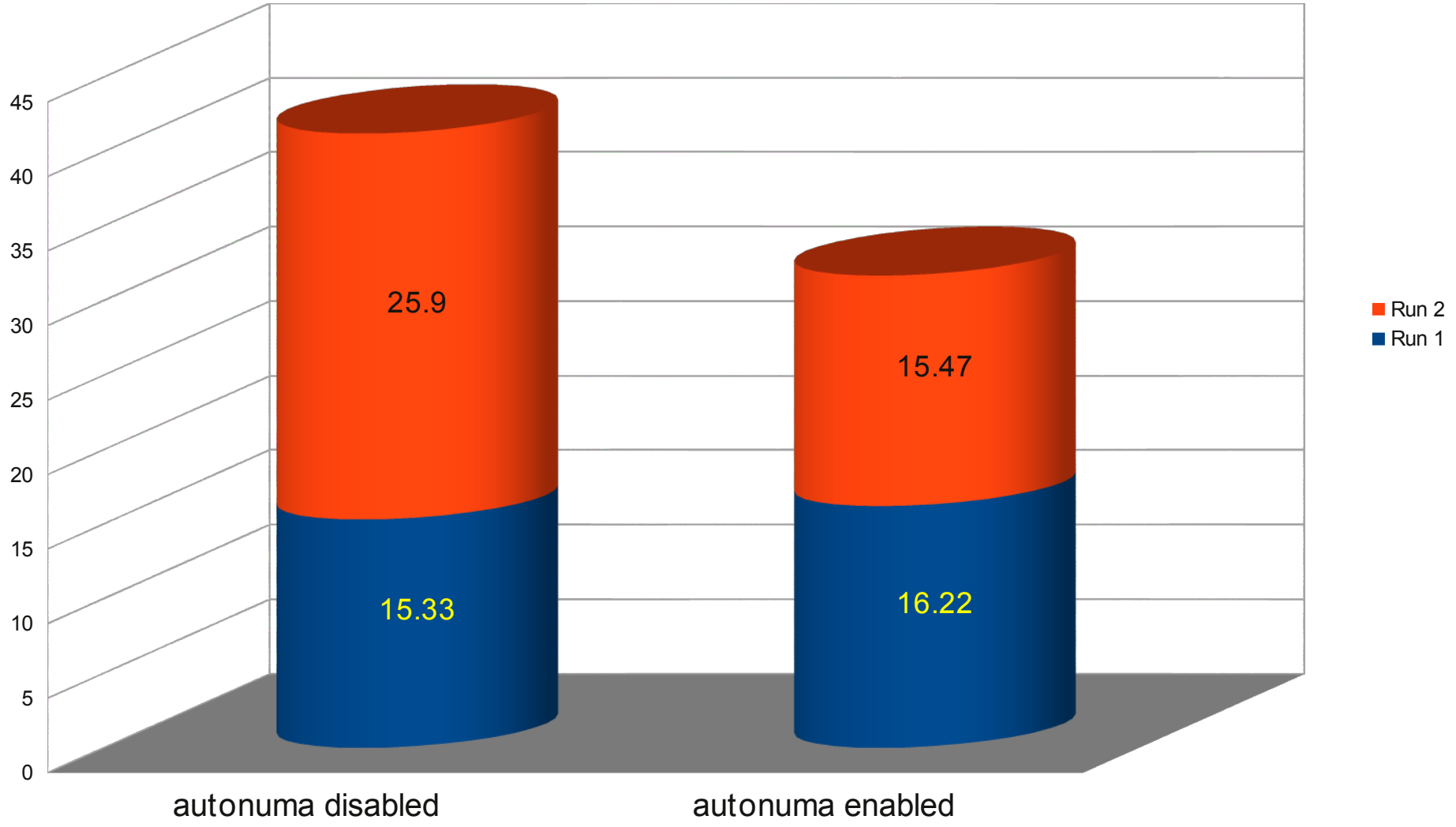
SPECjbb results 2 NUMA nodes, 8 CPUs per node, 16 CPUs total
THP enabled, no virt



Copyright © 2012 Red Hat Inc.

redhat

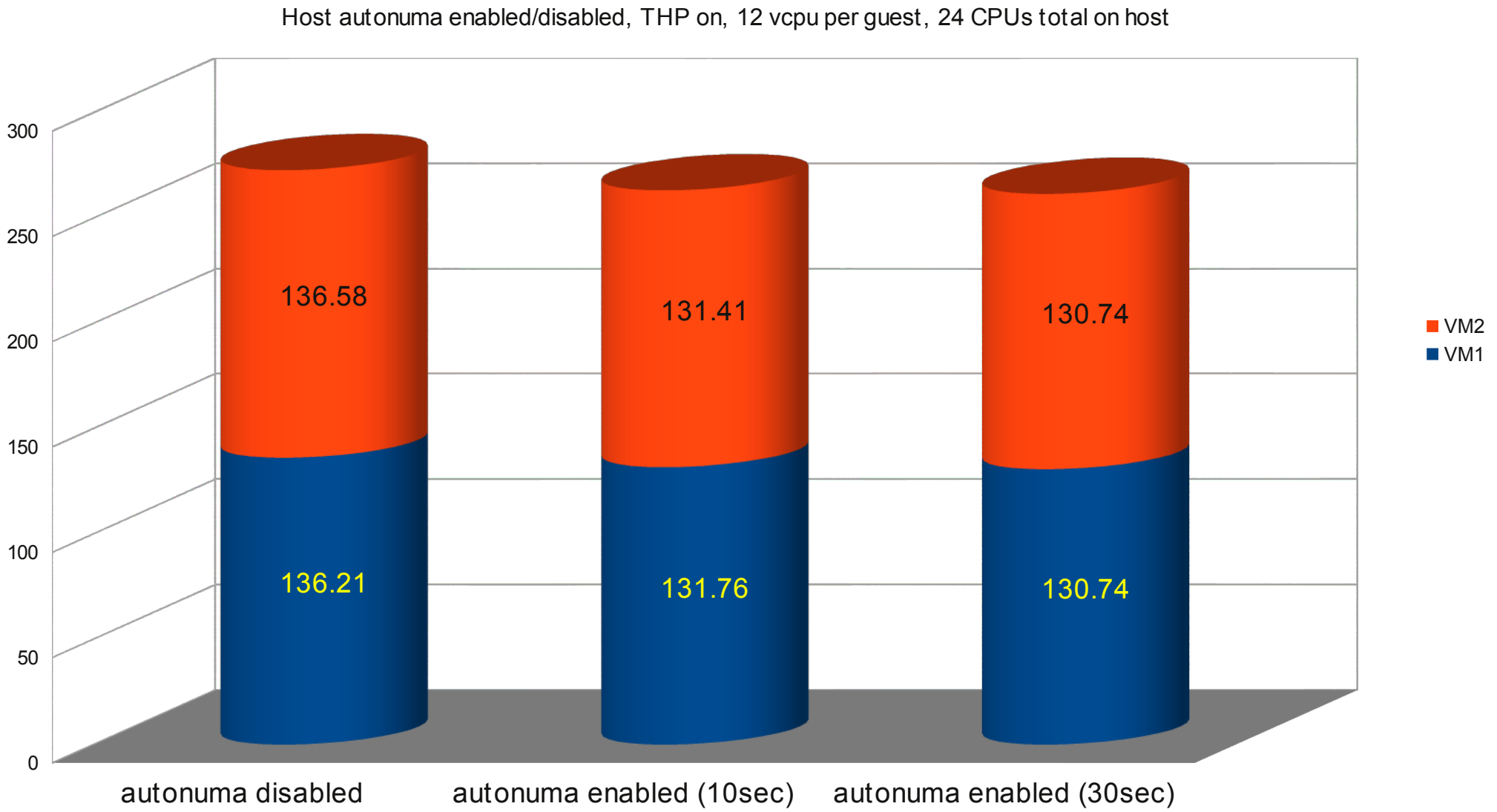
Virt guest "memhog -r100 1g" (autonuma includes 1 knuma_scand pass every 10 sec)
KVM host autonuma enabled/disabled, THP enabled
Guest VM fits in one host NUMA node



Copyright © 2012 Red Hat Inc.

redhat

kernel build -j16 in parallel in 2 KVM (both in tmpfs, in a loop started in sync)
Both guest VM fits in one host NUMA node
autonuma/knuma_scand/scan_sleep_pass_millisecs = 5000 | 15000 (10sec | 30sec)



Copyright © 2012 Red Hat Inc.

TODO: THP native migration

- THP native migration
  - SPECjbb results with khugepaged boosted shows the main bottleneck left is lack of THP native migration:
    - One copy in migration
    - One copy in khugepaged to rebuild the hugepage
  - Once this feature is added, AutoNUMA should perform even closer to numactl than it does now with khugepaged boosted (3rd column for every SPECjbb pass).
  - Urgent

Copyright © 2012 Red Hat Inc.

redhat.

TODO: struct page

- Allocate the 24 bytes per page only when booted on NUMA hardware like memcg does it:
  - Tricky with the different direct mapping implementation
  - At build time struct page is already not enlarged if CONFIG_AUTONUMA=n
  - Not too urgent
- No need to shrink the 24 bytes. Perhaps we could save 8 bytes, by crippling the list implementation (like forbidding list-deletion operations), but list-deletion (for migration cancellation when false sharing is detected) is already used.

Copyright © 2012 Red Hat Inc.

redhat.

TODO: document sched/numa.c

- Asked by Hillf as urgent item.
- Less urgent: write proper high level documentation to put in Documentation/vm/autonuma.txt .

Copyright © 2012 Red Hat Inc.

redhat.