

Hi,

This patch aims at allowing userspace to recreate the most important contents of /proc/stat per-cgroup. It exports the data needed for it from the guts of the scheduler, and then anyone can parse it and present it to a container in a meaningful way. Again, the kernel won't get involved in this directly.

Part of it will come from the cpu cgroup. Another part, from the cpuacct cgroup. Data is exported in cgroup files stat_percpu. They are just like the normal stat files, but with a cpuXXX value before the actual data field. As so, they are also extensible. So if anyone wants to give a shot at values I am currently ignoring (as iowait) in the future, we at least won't have a format problem.

Let me know what you think.

Glauber Costa (5):

- measure exec_clock for rt sched entities
- account guest time per-cgroup as well.
- record nr_switches per task_group
- expose fine-grained per-cpu data for cpuacct stats
- expose per-taskgroup schedstats in cgroup

kernel/sched/core.c | 214 ++++++
kernel/sched/fair.c | 27 +++++-
kernel/sched/rt.c | 5 +
kernel/sched/sched.h | 6 ++
4 files changed, 245 insertions(+), 7 deletions(-)

--

1.7.7.6
