Contribution ID: **309** Type: **not specified**

How many CPUs do I have? ...and other perplexing questions containers must answer

Thursday, 19 September 2024 17:15 (20 minutes)

One question applications running in containers often ask is: how many CPUs do I have access to? They want to know, e.g., how many threads they can run in parallel for their threadpool size, or the number of thread-local memory arenas.

The kernel offers many endpoints to query this information. There is /proc/cpuinfo, /proc/stat, sched_getaffinity(), sysinfo(), the cpuset cgroup hierarchy's cpuset.cpus.effective, the isolcpus kernel command line parameter, /sys/devices/system/cpu/online. Further, libcs offer divergent implementations of sysconf(_SC_NPROCESORS_ONLIN). As a bonus, the kernel scheduler may be configured to limit resources using cpu "shares" or cpu quotas, so a task may be able to run on all cores, but have some kind of rate limit that is not reflected in the physical cores the system is allowed to run on. Or, if SCHED_EXT lands as expected in 6.11, this whole concept will be configurable in userspace.

This discussion is an extension of one that began at FOSDEM'24 1, where we proposed a rust library for users to link against that would contain this information. In the hallway track of that conference, we ended up talking with systemd folks, who asked for an RFE2 for a /var/link interface so that this could be determined by IPC instead of by library.

There are advantages and drawbacks to both approaches. A library will require all language runtimes to modify their builds and add dependencies, which will be a tough sell. An IPC mechanism will require containers to be running this code, or the host running the code. In the IPC case, is some question about the container's cgns and resolving this stuff below the delegation boundary.

The goal of this talk will be to come away with a decision on a path forward.

Primary author: ANDERSEN, Tycho (Netflix)

Presenter: ANDERSEN, Tycho (Netflix)

Session Classification: Containers and checkpoint/restore MC

Track Classification: Containers and checkpoint/restore MC