CPU affinity, CPU sets, containers, etc.

Problem description

It may happen that a Grid job allocated to e.g. a single slot/single core runs very CPU intensive commands (for example, *make -j9*). If worker nodes are shared across multiple jobs/users, these commands may penalize other jobs running on the same WN.

Possible solutions

VO wrapper script

A job wrapper could use tasksets to assign processor affinities. For example, a pilot or job wrapper could issue the statement `os.system('taskset -p 0x00000001 %s' % os.getpid())` so that the process is limited to CPU0.

However, this is a problem that should be handled at the LRMS level.

Tasksets at the LRMS level

Rather than manually maintaining a list of CPUs bound to applications (or jobs), it is much simpler to use them when there is specific LRMS support for pinning jobs to CPUs. Platform LSF supports pinning with a variety of options [3].

CPU sets

CPU sets [1] are logical, hierarchical groupings of CPUs and units of memory; these groups can be bound jobs/applications, so that these are constrained in the resources they may use. CPU sets are natively supported in RHEL 6. The support for CPU sets in LRMS' needs to be investigated and tested, but they could simplify management of both single- and multi-core processes. PBS/Torque support them [2].

CPU sets are only applicable to WLCG, however, when jobs can run on RHEL 6 WNs.

lxc, Linux Containers

A completely different approach to process isolations can be obtained through Linux Containers [4], which uses the *cgroup* filesystem (available in RHEL 6) and implements resource isolation creating "containers", i.e. virtual systems. Like CPU sets, however, lxc require RHEL 6 WNs and should be integrated in LRMS support; some documentation on using lxc with Condor and PBS is available in [5].

Notes

[1] cpuset, RedHat 6 documentation,
http://docs.redhat.com/docs/en-US/Red_Hat_Enterprise_Linux/6/html/Resource_Management_Guide/sec-cpuset.html

[2] Torque Linux Cpuset Support,
http://www.adaptivecomputing.com/resources/docs/torque/2-5-9/3.5linuxcpusets.php

[3] Processor binding for LSF job processes,
http://www.ccs.miami.edu/hpc/lsf/7.0.6/admin/scalability_performance.html#wp4653559

[4] lxc Linux Containers, http://lxc.sourceforge.net/

[5] Brian Bockelman, Mixing SL5 and SL6 with chroot, GDB January 2012,
http://indico.cern.ch/getFile.py/access?sessionId=5&resId=0&materialId=0&confId=155064

-- DavideSalomoni - 03-Feb-2012

This topic: LCG > WMTEGCPUAffinity
Topic revision: r1 - 2012-02-03 - DavideSalomoni