

Arvados - Feature #8016

[Crunch2] Report statistics

12/15/2015 08:45 PM - Peter Amstutz

Status: Resolved	Start date: 06/30/2016
Priority: Normal	Due date:
Assigned To: Tom Clegg	% Done: 100%
Category:	Estimated time: 0.00 hour
Target version: 2016-07-06 sprint	
Description	
Subtasks:	
Task # 9481: Refactor crunchstat as a module	Resolved
Task # 9482: Use crunchstat in crunch-run to report container resource usage	Resolved
Task # 9484: Review 8016-crunchrun-crunchstat	Resolved

Associated revisions

Revision 257d6025 - 07/05/2016 05:28 PM - Tom Clegg

Merge branch '8016-crunchrun-crunchstat'

closes #8016

Revision 6a73335d - 07/05/2016 06:47 PM - Tom Clegg

8016: Rename Poll to PollPeriod. Amends c63c699aa9948f6a672536ba08e664471fb0d654.

refs #8016

History

#1 - 06/14/2016 06:44 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#2 - 06/14/2016 06:45 PM - Tom Clegg

Refactor crunchstat as a module, so crunch-run can start up a goroutine that monitors the correct cgroup (based on what docker client library tells us) and writes the appropriate log events/collections.

#3 - 06/22/2016 08:01 PM - Tom Clegg

- Target version set to 2016-07-06 sprint

#4 - 06/22/2016 08:02 PM - Tom Clegg

- Assigned To changed from Peter Amstutz to Tom Clegg

#5 - 06/22/2016 08:02 PM - Tom Clegg

- Story points set to 1.0

#6 - 06/30/2016 01:12 PM - Tom Clegg

- Status changed from New to In Progress

#7 - 07/01/2016 01:59 PM - Tom Clegg

8016-crunchrun-crunchstat @ [c9e6477](#)

#8 - 07/04/2016 05:55 PM - Radhika Chippada

lib/crunchstat/crunchstat.go

- line 314: missing err check
- In doCPUStats, need any consideration for the case when getCPUCount return 0 (due to errors caught)?

- How / where is lastCPUSample.hasData being set?
- "If CID is empty, wait for it to appear in CIDFile. Return true if we get it before r.done indicates someone called Stop" => "If CID is empty, wait for it to appear in CIDFile. Return true if we get it before r.done **which** indicates someone called Stop"? Similar comment at line 409.
- Should we be returning true if CIDFile is empty and CID is not available here?

```
func (r *Reporter) waitForCIDFile() bool {
    if r.CID != "" || r.CIDFile == "" {
        return true
    }
}
```

- In Reporter struct, can we call Poll as PollInterval?

services/crunch-run/crunchrun.go

- "resource usage statistics reporting period" => can you please expand this a little? I am understanding this as resource usage stats will be reported at regular intervals given?

#9 - 07/04/2016 09:26 PM - Tom Clegg

Radhika Chippada wrote:

lib/crunchstat/crunchstat.go

- line 314: missing err check

Fixed. (The "return zero" outcome is the same but better to do it explicitly.)

- In doCPUStats, need any consideration for the case when getCPUCount return 0 (due to errors caught)?

We'll report "0 cpus" and the downstream analysis will have to cope with it.

I think this is preferable to dropping the rest of the information for this sample period. It probably means either that method never works on this system, or something terrible has happened like we're running out of file descriptors (and the times when the worker host is having trouble are probably the times when we are most eager to save all the information we can).

- How / where is lastCPUSample.hasData being set?

This was set by nextSample := cpuSample{true, ...}. I've fixed that to use named fields, so it's less mysterious.

```
-     nextSample := cpuSample{true, time.Now(), 0, 0, r.getCPUCount()}
-     var userTicks, sysTicks int64
-     fmt.Sscanf(string(b), "user %d\nsystem %d", &userTicks, &sysTicks)
-     userHz := float64(C.sysconf(C._SC_CLK_TCK))
-     nextSample.user = float64(userTicks) / userHz
-     nextSample.sys = float64(sysTicks) / userHz
+     nextSample := cpuSample{
+         hasData:    true,
+         sampleTime: time.Now(),
+         user:        float64(userTicks) / userHz,
+         sys:         float64(sysTicks) / userHz,
+         cpus:        r.getCPUCount(),
+     }
```

- "If CID is empty, wait for it to appear in CIDFile. Return true if we get it before r.done indicates someone called Stop" => "If CID is empty, wait for it to appear in CIDFile. Return true if we get it before r.done **which** indicates someone called Stop"? Similar comment at line 409.

This was supposed to read as "before r.done indicates (by being closed) that someone called Stop" but I see how it doesn't scan well. Updated to "...before we learn (via r.done) that someone called Stop".

- Should we be returning true if CIDFile is empty and CID is not available here?
[...]

Yes. If CIDFile is empty, there's no CIDFile to wait for (returning true just means nobody called Stop before we returned).

If both are empty, we're going to collect host stats instead of container stats.

- In Reporter struct, can we call Poll as PollInterval?

Sure. I went with PollPeriod since it's more suggestive (in case anyone cares) that we're doing "poll once every X seconds", rather than "poll, then wait X seconds, then poll again".

services/crunch-run/crunchrun.go

- "resource usage statistics reporting period" => can you please expand this a little? I am understanding this as resource usage stats will be reported at regular intervals given?

Changed to "sampling period for periodic resource usage reporting" ... better?

Now at [cf1f493](#)

#10 - 07/05/2016 04:20 PM - Radhika Chippada

LGTM

#11 - 07/05/2016 05:35 PM - Tom Clegg

- *Status changed from In Progress to Resolved*

Applied in changeset arvados|commit:257d60253246952b435cea23b1912af80ea2c6d6.